

DOMINADOS PELOS NÚMEROS

do Facebook e
Google às fake news

os algoritmos que
controlam nossa vida

DAVID SUMPTER



DAVID SUMPTER

DOMINADOS PELOS NÚMEROS

do Facebook e
Google às fake news
—
os algoritmos que
controlam nossa vida

Tradução

Anna Maria Sotero e Marcello Neto

1ª edição

B
BERTRAND BRASIL

Rio de Janeiro | 2019

Copyright © 2018, David Sumpter

Esta tradução de *Outnumbered* foi publicada pela Bertrand Brasil em acordo com a Bloomsbury Publishing Inc. Todos dos direitos reservados.

Título original: *Outnumbered*

Texto revisado segundo o novo Acordo Ortográfico da Língua Portuguesa

2019

Impresso no Brasil

Printed in Brazil

CIP-BRASIL. CATALOGAÇÃO NA PUBLICAÇÃO
SINDICATO NACIONAL DOS EDITORES DE LIVROS, RJ

Sumpter, David

S953d Dominados pelos números [recurso eletrônico] : do facebook e google às fake news : os algoritmos que controlam nossa vida / David Sumpter ; tradução Anna Maria Sotero, Marcello Neto. - 1. ed. - Rio de Janeiro : Bertrand Brasil, 2019.
: il.

Tradução de: *Outnumbered*

Formato: epub

Requisitos do sistema: adobe digital editions

Modo de acesso: world wide web

Inclui índice

ISBN 978-85-286-2398-7 (recurso eletrônico)

1. Modelos matemáticos - Aspectos Sociais. 2. Comportamento humano - Modelos matemáticos. 3. Redes sociais. 4. Livros eletrônicos. I. Sotero, Anna Maria. II. Neto, Marcello. III. Título.

19-55138

CDD: 303.483

CDU: 316.422.44

Leandra Felix da Cruz - Bibliotecária - CRB-7/6135

Todos os direitos reservados. Não é permitida a reprodução total ou parcial desta obra, por quaisquer meios, sem a prévia autorização por escrito da Editora.

Direitos exclusivos de publicação em língua portuguesa somente para o Brasil adquiridos pela:
EDITORA BERTRAND BRASIL LTDA.

Rua Argentina, 171 – 3º andar – São Cristóvão

20921-380 – Rio de Janeiro – RJ

Tel.: (21) 2585-2000 – Fax: (21) 2585-2084

Atendimento e venda direta ao leitor:

mdireto@record.com.br ou (21) 2585-2002

Sumário

PARTE 1: ANALISANDO-NOS

Capítulo 1: Procurando Banksy

Capítulo 2: Gerando ruído

Capítulo 3: Os principais elementos da amizade

Capítulo 4: Suas 100 dimensões

Capítulo 5: A hiperbolitização de Cambridge

Capítulo 6: Impossivelmente imparcial

Capítulo 7: Os alquimistas dos dados

PARTE 2: INFLUENCIANDO-NOS

Capítulo 8: Nate Silver *versus* o restante de nós

Capítulo 9: Nós “também gostamos” da internet

Capítulo 10: O concurso de popularidade

Capítulo 11: Borbulhando

Capítulo 12: O futebol importa

Capítulo 13: Quem lê fake news?

PARTE 3: IMITANDO-NOS

Capítulo 14: Aprendendo a ser sexista

Capítulo 15: Apenas pensamento entre decimal

Capítulo 16: Arrase no *Space Invaders*

Capítulo 17: O cérebro bacteriano

Capítulo 18: De volta à realidade

Notas

Agradecimientos

Índice

PARTE 1

ANALISANDO-NOS

PROCURANDO BANKSY

Em março de 2016, três pesquisadores de Londres e um criminologista do Texas publicaram um artigo no *Journal of Spatial Science*. Os métodos apresentados foram práticos e eruditos, mas o artigo não era um exercício acadêmico abstrato. O título deixava o objetivo claro: “Marcando Banksy: usando perfis geográficos para investigar um mistério da arte moderna”. A matemática estava sendo usada para rastrear o grafiteiro mais famoso do mundo.

Os pesquisadores usaram o website de Banksy para identificar a localização de seu trabalho de rua. Então, visitaram sistematicamente cada um de seus trabalhos, tanto em Londres quanto em sua cidade natal, Bristol, com um gravador de GPS. Com os dados coletados, criaram um mapa de calor no qual áreas mais quentes mostravam onde era mais provável que Banksy houvesse morado, supondo que ele tivesse criado seu trabalho perto de casa.

O ponto mais quente do geoperfil de Londres ficava a apenas quinhentos metros do antigo endereço da namorada de um homem que já havia sido apontado como Banksy. No mapa de Bristol, a cor estava mais quente ao redor da casa onde esta mesma pessoa vivia e do campo de futebol do time em que jogava. O artigo concluiu que a pessoa deste geoperfil era muito provavelmente Banksy. Ao ler o artigo, minha primeira reação foi uma

mistura de interesse e inveja que a maioria dos acadêmicos sente quando seus colegas fazem algo engenhoso. Essa foi uma escolha inteligente de aplicação. Exatamente o tipo de matemática aplicada que eu ambiciono fazer: bem executada, com um algo a mais que captura a imaginação. Gostaria de ter feito isso eu mesmo.

Mas, à medida que fui lendo, comecei a ficar um pouco irritado. Eu gosto do Banksy. Tenho o livro com seus desenhos e citações irreverentes na minha mesa de centro. Já perambulei por ruelas de cidades procurando por seus murais. Dei risadas com o vídeo do fracasso de vendas de seu valioso trabalho em uma barraquinha montada por ele no Central Park em Nova York. Suas criações na Cisjordânia e nos campos de imigrantes de Calais fizeram com que eu me sentisse desconfortável em relação à minha posição privilegiada. Eu não quero alguns acadêmicos emocionalmente desconectados usando seus algoritmos para me dizer quem é Banksy. O grande lance de Banksy é que ele trabalha sorrateiramente durante a noite e, na luz do amanhecer, sua arte revela as hipocrisias de nossa sociedade.

A matemática está destruindo a arte: uma estatística lógica e fria está perseguindo, pelas ruelas de Londres, combatentes da liberdade encapuzados. Isso não está certo. A polícia e a imprensa sensacionalista é que deveriam estar procurando Banksy, não acadêmicos liberais. Quem esses espertinhos pensam que são?

Li o artigo de Banksy poucas semanas antes da data prevista para a publicação de meu livro *Soccermaths*. Meu objetivo ao escrever um livro sobre futebol foi proporcionar aos leitores uma viagem matemática ao universo deste belo jogo. Quis mostrar que a estrutura e os padrões inerentes ao futebol estavam cheios de matemática.

Quando o livro foi lançado, tal ideia causou bastante interesse na mídia, e todo dia me convidavam para dar entrevistas. Na maioria das vezes, os jornalistas estavam tão fascinados quanto eu pela matemática do futebol, mas havia uma pergunta incômoda que sempre vinha à baila. Uma pergunta para a qual, segundo eles, os leitores iriam querer saber a resposta: “Você acha que os números estão tirando a paixão do jogo?”

“Claro que não!”, respondi com indignação. Expliquei que sempre haverá lugar tanto para o pensamento lógico quanto para a paixão no grande templo que é o futebol.

Mas as revelações matemáticas não desmistificaram um pouco a arte de Banksy? Eu não estava, cinicamente, usando o futebol da mesma forma? Talvez eu estivesse fazendo com os fãs de futebol exatamente a mesma coisa que os estatísticos espaciais estavam fazendo com os grafiteiros de rua.

No final daquele mês, fui convidado para dar uma palestra sobre a matemática do futebol na sede londrina do Google. O evento foi organizado pela assessora de imprensa do livro, Rebecca, e nós dois estávamos ansiosos para conhecer as instalações de pesquisa do Google.

Não ficamos desapontados. Os escritórios ficavam num prédio bem equipado na Buckingham Palace Road, onde havia grandes estruturas de Lego no hall de entrada e geladeiras lotadas de bebidas saudáveis e superalimentos. Os “googlers”, como eles se referiam a si mesmos, claramente tinham muito orgulho de seu ambiente de trabalho.

Perguntei a alguns dos googlers no que a empresa estava envolvida naquele momento. Eu tinha ouvido falar dos carros autônomos, do Google Glass e das lentes de contato inteligentes, dos drones que entregam encomendas na nossa porta e da ideia de injetar nanopartículas no nosso corpo para detectar doenças, e eu queria saber mais sobre os rumores.

Mas os googlers foram cautelosos. Depois de uma série de matérias negativas sobre a empresa estar se tornando um centro excessivamente criativo de ideias insanas, a política era parar de divulgar para o mundo exterior informações demais sobre o que estava sendo desenvolvido. A chefe de projetos tecnológicos avançados daquela época, Regina Dugan, anteriormente tivera o mesmo cargo na Agência de Projetos de Pesquisa Avançada de Defesa (Darpa), do governo americano. Ela estabeleceu regras do tipo “necessidade de saber” para o compartilhamento de informações no Google. O departamento de pesquisa passou a consistir em pequenas unidades, cada uma trabalhando no seu próprio projeto e compartilhando ideias e dados internamente dentro dos grupos.¹

Depois de tanto perguntar, um dos googlers finalmente mencionou um projeto. “Ouvi que estamos usando a DeepMind para examinar diagnósticos médicos de insuficiência renal”, disse ele.

O plano era utilizar o aprendizado de máquina para encontrar padrões nas doenças renais que os médicos não perceberam. A DeepMind em questão é a filial do Google que programou um computador para se tornar o melhor jogador de Go do mundo e treinou um algoritmo para dominar o *Space Invaders* e outros jogos de fliperama antigos. Agora pesquisaria os registros médicos do NHS, o serviço nacional de saúde do Reino Unido, para tentar encontrar padrões na ocorrência de doenças. A DeepMind se tornaria um assistente computacional inteligente para os médicos.

Assim como quando li pela primeira vez o artigo sobre Banksy, senti novamente aquela excitante pontinha de desejo dos googlers, porém com um pouco de inveja de também ter a chance de descobrir doenças e melhorar os cuidados com a saúde, usando algoritmos. Imagine ter a verba e os dados para realizar um projeto como este e salvar vidas por meio da matemática.

Rebecca não ficou tão impressionada. “Não sei se eu gostaria que o Google tivesse acesso a todos os meus dados médicos”, disse. “É um pouco preocupante quando você imagina como eles podem vir a usá-los junto com outras informações pessoais.”

A reação dela me fez pensar no assunto novamente. Bancos de dados detalhados, tanto com informações de saúde como de estilo de vida, estão se acumulando mais rápido do que nunca. Enquanto o Google seguia regras rígidas sobre a proteção de dados, a ameaça ainda existia. No futuro, a sociedade pode exigir que conectemos nossas buscas, redes sociais e informações médicas, a fim de ter uma ideia completa de quem somos e por que ficamos doentes.

Não tivemos muito tempo para discutir os prós e contras da pesquisa médica orientada por dados antes da minha apresentação. E assim que comecei a falar sobre futebol, esqueci o assunto. Os googlers ficaram fascinados e fizeram muitas perguntas. Qual era a mais recente tecnologia de

ponta sobre câmeras para rastreamento? Os dirigentes do futebol poderiam ser substituídos por modelos de aprendizado de máquina que gradualmente melhorariam as táticas? Também houve perguntas técnicas sobre coleta de dados e futebol robótico.

Nenhum dos googlers me perguntou se eu achava que todos esses dados dilacerariam a alma do jogo. Provavelmente, eles ficariam muito felizes se pudessem conectar em todos os jogadores os equipamentos de monitoramento de saúde e de nutrição por 24 horas a fim de obter um perfil completo de suas condições físicas. Quanto mais dados, melhor.

Eu entendo os googlers, assim como entendo os estatísticos do caso Banksy. É muito legal ter o banco de dados de um paciente da NHS em seu computador ou ser capaz de rastrear “criminosos” usando estatística espacial. Em Londres, Berlim, Nova York, Califórnia, Estocolmo, Xangai e Tóquio, nerds da matemática, assim como nós, estão coletando e processando dados. Estamos desenvolvendo algoritmos para reconhecer rostos, entender idiomas e aprender nossas preferências musicais. Estamos construindo assistentes pessoais e chatbots (interfaces conversacionais) que o ajudam a consertar seu computador. Estamos prevendo os resultados de eleições e de competições esportivas. Estamos encontrando o par perfeito para pessoas solteiras ou ajudando-as a conferir todas as possibilidades disponíveis. Estamos tentando veicular as notícias mais relevantes para você no Facebook e no Twitter. Estamos nos certificando de que você encontre os melhores pacotes de viagens e voos promocionais. Nosso objetivo é usar os dados e algoritmos como uma força para o bem.

Mas é realmente assim tão simples? Os matemáticos estão tornando o mundo um lugar melhor? Minha reação ao desmascaramento de Banksy, a reação dos jornalistas aos meus modelos do livro *Soccermetrics* e a reação da Rebecca ao uso do banco de dados médicos pelo Google não são incomuns nem infundadas. Elas são bastante normais. Os algoritmos são usados em várias situações para nos ajudar a entender melhor o mundo. Mas será que realmente queremos entender melhor o mundo, se isso significar dissecar as coisas que amamos e perder nossa integridade pessoal? Os algoritmos que

desenvolvemos estão realizando as coisas que a sociedade deseja ou estão somente servindo aos interesses de alguns geeks e das empresas multinacionais para as quais eles trabalham? Será que existe o risco de os algoritmos assumirem o controle de tudo à medida que desenvolvemos uma inteligência artificial (IA) cada vez melhor? E de que a matemática comece a tomar decisões no nosso lugar?

A maneira por meio da qual o mundo real e a matemática interagem nunca é simples e direta. Todos nós, inclusive eu, algumas vezes caímos no erro de achar que a matemática é um dispositivo em que se gira uma manivela e os resultados são ejetados. Os matemáticos aplicados são treinados para enxergar o mundo em termos de um ciclo de modelagem. O ciclo começa quando consumidores do mundo real nos dão um problema que eles querem resolver, seja procurar Banksy ou desenvolver uma ferramenta de busca on-line. Nós pegamos nossa caixa de ferramentas matemáticas, ligamos nossos computadores, programamos nossos códigos e vemos se conseguimos obter respostas melhores. Implementamos um algoritmo e o fornecemos aos clientes que o solicitaram. Eles nos dão um feedback e o ciclo continua.

O girar da manivela e a modelagem em ciclo tornam os matemáticos desapegados. Os escritórios do Google e do Facebook, equipados com jogos e áreas de lazer internas, dão aos seus empregados superespertos a ilusão de que eles têm o controle total dos problemas de que estão tratando. O isolamento esplêndido dos departamentos universitários significa que não temos que confrontar nossas teorias com a realidade. Isso está errado. O mundo real tem problemas reais, e é nosso dever apresentar soluções reais. Cada problema é muito mais complexo e exige muito mais do que simplesmente alguns cálculos.

Nos meses que se seguiram à minha visita ao Google em maio de 2016, comecei a perceber um novo tipo de histórias matemáticas nos jornais. Uma incerteza estava se espalhando pela Europa e pelos Estados Unidos. A ferramenta de busca do Google estava gerando sugestões automáticas racistas; twitterbots estavam espalhando notícias falsas; Stephen Hawking

estava preocupado com a inteligência artificial; grupos de extrema direita estavam vivendo em bolhas-filtros criadas algoritmicamente; o Facebook estava avaliando a personalidade dos usuários e os resultados estavam sendo manipulados para selecionar eleitores. As histórias sobre os perigos dos algoritmos acumulavam-se uma após a outra. Até mesmo a capacidade dos matemáticos em fazer previsões foi questionada, uma vez que os modelos estatísticos erraram a respeito do Brexit e de Trump.

Histórias sobre a matemática do futebol, do amor, dos casamentos, dos grafites e de outras coisas divertidas foram, subitamente, substituídas por uma matemática do sexismo, do ódio, da distopia e de erros absurdos em cálculos de pesquisas de opinião.

Quando reli o artigo científico sobre Banksy, com um pouco mais de cuidado dessa vez, descobri que muito pouca evidência nova foi apresentada sobre sua identidade. Embora os pesquisadores tenham mapeado a posição exata de 140 obras de arte, eles só investigaram o endereço de um único suspeito. Oito anos antes, o suspeito em questão já havia sido identificado no *Daily Mail* como sendo o verdadeiro Banksy.² O jornal estabeleceu que ele vinha de uma família de classe média e que não era um herói da classe trabalhadora, como esperamos que grafiteiros sejam.

Um dos coautores do artigo científico, Steve Le Comber, foi sincero com a BBC sobre o porquê de eles terem focado no suspeito do *Daily Mail*. Ele disse: “Em pouco tempo ficou evidente que só existe um único suspeito sério e todo mundo sabe quem ele é. Se você pesquisar no Google por Banksy e [nome do suspeito], você obterá cerca de 43.500 resultados”.³

Muito antes de os matemáticos aparecerem, a internet achou que sabia a verdadeira identidade de Banksy. O que os pesquisadores fizeram foi associar números àquele conhecimento, mas não é muito claro o que esses números significam. Os cientistas só haviam testado um suspeito em um caso. O artigo ilustrou os métodos, mas estava longe de ser uma prova conclusiva de que esses métodos funcionaram de fato.

A mídia não estava muito preocupada com as limitações do estudo. Uma história sensacionalista do *Daily Mail* tinha se tornado uma notícia séria,

coberta pelo *Guardian*, pelo *Economist* e pela BBC. A matemática deu legitimidade a um boato. E o estudo contribuiu para a crença de que a tarefa de encontrar “criminosos” poderia ser realizada por um algoritmo.

Mudemos nosso cenário para um tribunal. Imagine que, em vez de estar sendo acusado de nos entreter com sua reverenciada arte de rua, Banksy seja um muçulmano acusado de cobrir as paredes da cidade de Birmingham com propaganda pró-Estado Islâmico. Vamos imaginar ainda que a polícia tenha feito uma pesquisa de antecedentes e tenha descoberto que a pichação começou quando o suspeito se mudou de Islamabad para Birmingham. Mas eles não podem usar esse fato na corte, pois não há evidências. Então, o que a polícia faz? Bem, eles chamam os matemáticos. Usando seus algoritmos, os peritos estatísticos da polícia preveem com 65,2% de certeza que uma casa em particular pertence ao Banksy islâmico e o esquadrão antiterrorismo é chamado. Uma semana depois, sob a Lei de Prevenção ao Terrorismo, o Banksy islâmico é posto em prisão domiciliar.

Esse cenário não é muito diferente daquele que Steve e seus colegas conceberam usando os resultados de suas pesquisas. No artigo, eles escrevem que encontrar Banksy “endossa sugestões anteriores de que a análise de atos leves relacionados ao terrorismo (por ex., pichação) poderia ser usada para ajudar a localizar bases terroristas antes que incidentes mais sérios aconteçam”. Com o respaldo matemático em mãos, o Banksy islâmico é acusado e condenado. O que era antes uma fraca evidência circunstancial é agora prova estatística.

Esse é apenas o começo. Após o sucesso na identificação do Banksy islâmico, empresas privadas começaram a disputar contratos para prover aconselhamento estatístico para a força policial. Depois de conseguir seu primeiro contrato, o Google alimenta a DeepMind com todos os registros criminais para identificar terroristas em potencial. Poucos anos depois, o governo, apoiado pelo público em geral, introduz a medida de “senso comum”, de integrar nossa busca na internet com o banco de dados de registros policiais do Google. O resultado é a criação de “autoridades com inteligência artificial” que podem usar nossas buscas e os dados de

localização para ponderar nossa motivação e comportamento futuro. À Autoridade IA é atribuída uma força de ataque designada para executar invasões noturnas aos supostos terroristas em potencial. Num ritmo alarmante, começa o futuro sombrio da matemática.

Ainda estamos nas primeiras páginas e a matemática não está apenas sendo uma desmancha-prazeres, está comprometendo nossa integridade pessoal, está fornecendo legitimidade à imprensa sensacionalista, está acusando os cidadãos de Birmingham de cometer atos terroristas, está compilando uma imensa quantidade de dados dentro de corporações gigantes que não são responsabilizáveis e está construindo um supercérebro para monitorar nosso comportamento.

Quão sérias são essas questões e quão realistas são esses cenários? Decidi descobrir do único modo que sabia: analisando os dados, calculando as estatísticas e fazendo as contas.

GERANDO RUÍDO

Depois que Banksy foi desmascarado matematicamente, notei que, de certo modo, eu não tinha acompanhado a real importância da mudança que os algoritmos estavam causando na nossa sociedade. Não me entenda mal. Certamente não deixei de acompanhar o desenvolvimento da matemática. Aprendizado de máquina, modelos estatísticos e inteligência artificial são tópicos que pesquiso e discuto ativamente com meus colegas todos os dias. Eu leio os artigos mais recentes e me mantenho atualizado com os avanços mais importantes. Mas estava me concentrando no lado científico das coisas: analisando como os algoritmos funcionam do ponto de vista abstrato. Falhei ao não pensar nas consequências de suas aplicações. Não pensei em como as ferramentas que estava ajudando a desenvolver poderiam mudar a sociedade.

Eu não era o único matemático a enfrentar essa revelação, e comparado à minha preocupação frívola sobre a identidade de Banksy ser revelada, alguns de meus colegas descobriram coisas realmente preocupantes. Quase no fim de 2016, a matemática Cathy O’Neil publicou o livro *Weapons of Math Destruction*, no qual documenta os usos indevidos de algoritmos em tudo, desde avaliação de professores e propaganda on-line de cursos universitários a fornecimento de crédito privado e previsões de reincidências criminais.¹ Suas conclusões eram assustadoras: os algoritmos

estavam nos julgando de maneira arbitrária, frequentemente baseados em pressuposições dúbias e dados imprecisos.

Um ano antes, Frank Pasquale, um professor de direito da Universidade de Maryland, publicara o livro *The Black Box Society*. Ele argumentou que, enquanto nossas vidas privadas estavam se tornando cada vez mais expostas — uma vez que compartilhamos detalhes de nosso cotidiano, nossas aspirações, nossos movimentos e nossas vidas sociais on-line —, as ferramentas usadas por Wall Street e pelas empresas do Vale do Silício para analisar nossos dados eram guardadas a sete chaves. Essas caixas-pretas estavam influenciando a informação que víamos e nos avaliando, ao mesmo tempo que nos mantinham incapazes de descobrir como esses algoritmos funcionam.

Encontrei pela internet um grupo informalmente organizado de cientistas de dados que estava reagindo a essas ameaças e analisando como os algoritmos eram aplicados na sociedade.

As preocupações mais imediatas desses ativistas giravam em torno da transparência e da capacidade de influenciar. Enquanto você está on-line, o Google coleta informações dos sites que você visita e usa esses dados para decidir que anúncios lhe mostrar. Faça uma pesquisa sobre a Espanha e nos próximos dias você verá opções de viagens para lá. Procure por futebol e mais sites de apostas aparecerão na sua tela. Procure por links sobre os perigos da caixa-preta dos algoritmos e lhe oferecerão uma assinatura do *New York Times*.

Com o tempo, o Google constrói um quadro de seus interesses e os classifica. É simples descobrir o que ele deduziu a seu respeito usando as “configurações de anúncios” na sua conta do Google.² Quando acessei essas configurações, descobri que o Google sabe bastante coisa a meu respeito: futebol, política, comunidades on-line e atividades ao ar livre estão todos corretamente identificados como coisas que me interessam. Mas alguns outros tópicos sugeridos eram um tanto incorretos: futebol americano e ciclismo são dois esportes de que o Google acha que eu gosto, mas não tenho, de fato, nenhum interesse neles. Senti que eu precisava dar um jeito

nisso. Em configurações de anúncios, cliquei no X ao lado dos esportes que não me interessam e, então, adicionei matemática à lista.

Na Universidade de Carnegie Mellon, na Pensilvânia, Estados Unidos, o aluno de doutorado Amit Datta e seus colegas conduziram uma série de experimentos para medir como exatamente o Google nos classifica. Eles projetaram uma ferramenta automática que cria “agentes” do Google que abrem páginas da internet com configurações predefinidas. Depois, esses agentes visitaram sites relacionados a assuntos específicos, e os pesquisadores olharam tanto para os anúncios que eram exibidos para os agentes quanto para as mudanças em suas configurações de anúncios. Quando os agentes navegaram em sites relacionados ao abuso de drogas, eram mostrados anúncios sobre clínicas de reabilitação. Do mesmo modo, agentes navegando em sites associados à deficiência física tinham uma chance maior de receber anúncios sobre cadeiras de roda. No entanto, o Google não é completamente honesto conosco. Em nenhum momento, as configurações de anúncios dos agentes foram atualizadas para apresentar ao usuário as conclusões que o algoritmo do Google tinha tirado a seu respeito. Mesmo quando usamos nossas configurações para informar ao Google quais anúncios queremos e quais não queremos ver, ele toma suas próprias decisões sobre o que nos mostrar.

Alguns leitores podem querer saber que o Google não mudou a propaganda para agentes que visitaram sites de conteúdo adulto. Quando perguntei a Amit se isso significava que usuários podem pesquisar livremente sobre pornografia sem aumentar a probabilidade de um anúncio inapropriado aparecer em suas telas em algum outro momento, ele aconselhou cuidado: “O Google pode mudar sua propaganda em outros sites que não navegamos. Então, anúncios inapropriados do Google podem surgir em outros sites.”

Todas as grandes empresas de serviços da internet — incluindo Google, Yahoo, Facebook, Microsoft e Apple — constroem um quadro personalizado de nossos interesses e o utilizam para decidir que anúncios nos mostrar. Esses serviços são transparentes até certo nível, permitindo aos usuários

fazer uma avaliação de suas configurações. É do interesse dessas empresas nos perguntar se elas entenderam corretamente nossas preferências. Mas, com certeza, elas não nos dizem tudo o que sabem a nosso respeito.

Angela Grammatas, que trabalha como programadora em marketing analítico, enfatizou que não há muita dúvida de que o redirecionamento — que é o termo técnico para a propaganda on-line que usa buscas recentes para escolher quais produtos mostrar para os usuários — é altamente eficaz. Ela me contou como a campanha “SoupTube”, da Campbell, usou o sistema Vagon do Google para mostrar aos usuários versões específicas de um anúncio que melhor correspondia aos seus interesses. De acordo com o Google, a campanha gerou um aumento de 55% nas vendas.³

Angela me disse: “O Google é relativamente benigno, mas o poder do botão ‘curtir’ do Facebook para selecionar anúncios é assustador. Sua ‘curtida’ fornece bastante insight sobre quem é você.” O que mais preocupou Angela foi uma mudança na lei nos Estados Unidos permitindo a provedores de serviço de internet (PSI) — as companhias de telecomunicação que fornecem internet residencial — arquivar e usar o histórico de busca de seus clientes. Diferentemente do Google e do Facebook, PSIs oferecem pouca ou nenhuma transparência sobre a informação que acumulam sobre você. Eles podem, potencialmente, vincular o histórico de navegação ao seu endereço de casa e compartilhar seus dados com anunciantes terceirizados.

Angela estava tão preocupada com essa mudança na lei que criou um plug-in para navegadores de internet que evitam que PSIs, ou quem quer que seja, acumulem informações úteis sobre seus clientes. Ela chamou o plug-in de Noiszy, uma vez que sua função é, literalmente, gerar ruído na navegação. Enquanto ela navega pelos sites que lhe interessam, o Noiszy trabalha paralelamente, navegando de forma aleatória por quarenta grandes sites de notícias. Os PSIs não têm como saber quais sites lhe interessam e quais não lhe interessam. A mudança nos anúncios mostrados no navegador dela foi imediata. “De repente eu estava vendo anúncios da Fox News... uma grande mudança em relação à bolha de ‘mídia liberal’ em que eu vivia

anteriormente”, ela me contou. Angela, que é casada, também percebeu um grande número de anúncios sobre vestidos de noiva. Seu navegador não sabia mais quem ela era.

Achei a abordagem da Angela fascinante porque ela parecia dividida sobre a questão de como as empresas usam nossos dados. Seu trabalho diário era criar anúncios de redirecionamento eficazes. Claramente, ela era muito boa no que fazia e acreditava que estava ajudando as pessoas a acharem os produtos que elas queriam. Mas, em seu tempo livre, ela tinha criado um plug-in que derrotava exatamente esse tipo de campanha publicitária e tinha disponibilizado seu software gratuitamente para qualquer um que o quisesse.⁴ “Se todos usarmos o Noiszy, empresas e organizações perdem a habilidade de nos compreender”, escreveu ela na página associada ao plug-in. Ela me disse que seu objetivo era aumentar o entendimento sobre o assunto e debater como a propaganda on-line funcionava.

Apesar das contradições evidentes, de algum modo entendi a lógica de Angela. Sim, houve casos evidentes de discriminação que precisavam ser detectados e interrompidos. Certamente, algumas das propagandas direcionadas para empréstimos a curto prazo e qualificações “universitárias” duvidosas eram imorais.⁵ E, sim, algumas vezes nosso navegador de internet tira conclusões ligeiramente estranhas a nosso respeito. Mas, em geral, o efeito da repetição de uma propaganda é relativamente benigno, e a maioria de nós não se incomoda que nos mostrem alguns produtos que possam nos interessar. Angela está certa ao focar em nos ensinar como a propaganda moderna funciona. É nossa responsabilidade entender os algoritmos que tentam nos vender coisas e garantir que os PSIs respeitem nossos direitos.

No entanto, as conclusões que os algoritmos tiram a nosso respeito podem ser discriminatórias. Para investigar preconceitos de gênero, Amit e seus colegas inicializaram 500 agentes “masculinos” (que tiveram seu sexo configurado como sendo masculino) e 500 agentes “femininos”, que navegaram em sites predefinidos relacionados a empregos.⁶ Depois dessas sessões de navegação, eles analisaram os anúncios mostrados aos agentes.

Apesar dos históricos de navegação parecidos, os homens tiveram mais chance de receber um anúncio específico do site careerchange.com que tinha a chamada: “empregos com salários a partir de US\$ 200 mil — somente executivos”. As mulheres tiveram mais chances de receber ofertas para sites de recrutamento genérico. Esse tipo de discriminação é descarada e potencialmente ilegal.

O presidente da empresa que administra o careerchange.com, Waffles Pi Natusch, contou ao *Pittsburgh Post-Gazette* que não sabia como os anúncios tenderam tão fortemente na direção dos homens, mas admitiu que algumas preferências dos anúncios da empresa — executivos com experiência, acima de 45 anos e salário acima de US\$ 100.000 por ano — poderiam levar os algoritmos do Google a isso.⁷ Essa explicação foi curiosa, uma vez que a diferença entre os agentes era apenas o gênero, e não o salário nem a idade. Ou o algoritmo de propaganda do Google relacionou, direta ou indiretamente, homens com executivos de alto salário ou o careerchange.com, inadvertidamente, selecionou a opção de direcionar seus anúncios para homens.⁸

Foi nesse ponto que a investigação de Amit e seus colegas terminou. Ele me contou que, embora não tivessem recebido nenhuma resposta do Google quando publicaram o trabalho, a gigante da internet mudou sua interface de modo que eles não puderam mais rodar seus experimentos com agentes. A caixa-preta foi fechada para sempre.

Julia Angwin e seus colegas da ProPublica, uma redação sem fins lucrativos, abriram várias caixas-pretas nos últimos dois anos, com uma série de reportagens sobre mecanismos tendenciosos. Usando dados coletados de mais de 7.000 réus na Flórida, Julia mostrou que um dos algoritmos mais usados pelo sistema judicial dos Estados Unidos era tendencioso contra afro-americanos.⁹ Mesmo considerando o histórico criminal, a idade, o gênero e futuros delitos, eles mostraram que os afro-americanos tinham uma chance de 45% de serem incluídos, pelo algoritmo, na categoria de alto risco de cometer um crime.

Esse tipo de discriminação não está limitado ao sistema legal. Em outro estudo da ProPublica, Julia criou um anúncio no Facebook direcionado aos “compradores de primeira viagem” e às pessoas que “eram propensas a se mudar”, mas que excluía aqueles que tinham uma “afinidade étnica” do tipo “afro-americano”, “americano-asiático” ou “hispânico”. O Facebook aceitou e publicou o anúncio apesar de violar a Lei do Direito à Moradia dos Estados Unidos.¹⁰ Excluir certos grupos, mesmo com base em “afinidade” (que o Facebook mede examinando as páginas e as publicações com os quais os usuários interagem) em vez da raça propriamente dita, é discriminação.

Os jornalistas da ProPublica fazem parte de um grupo muito maior de jornalistas de dados e cientistas que investigam essas questões. Joy Buolamwini, uma estudante de pós-graduação do MIT (Massachusetts Institute of Technology), descobriu que o sistema de reconhecimento facial atual não conseguia reconhecer o rosto dela. Assim, ela começou a reunir um conjunto de rostos mais etnicamente diversificado que poderia ser usado para treinar e melhorar sistemas de reconhecimento futuros.¹¹ Jonathan Albright, da Universidade Elon, na Carolina do Norte, investigou os dados usados pela ferramenta de busca do Google para tentar entender por que seu preenchimento automático gerava previsões de pesquisa racistas e ofensivas¹² e Jenna Burrell, da Universidade de Berkeley, Califórnia, usou engenharia reversa para analisar o programa do filtro de spam de sua conta de e-mail para verificar se ele discriminava nigerianos explicitamente (neste caso, não discriminou).¹³

Esses pesquisadores — juntamente com Angela Grammatas, Amit Datta, Cathy O’Neil e muitos outros — estão determinados a monitorar os algoritmos criados pelos gigantes da internet e das indústrias de segurança. Eles compartilham seus dados e seus códigos abertamente em repositórios de dados on-line, de modo que outras pessoas possam baixá-los e descobrir como eles funcionam. Muitos deles realizam esses estudos em seu tempo livre, usando suas habilidades de programadores, acadêmicos e estatísticos para descobrir como o mundo está sendo reformulado pelos algoritmos.

Dissecar algoritmos pode não ter a mesma reputação da arte de Banksy, mas em comparação com a perspectiva restrita do futuro e com as unidades de pesquisa sigilosas que encontrei nas instalações londrinas do Google, fiquei muito impressionado com o modo como esses ativistas trabalharam e forneceram seus resultados para qualquer um usá-los.

Esse movimento estava surtindo efeito. O Facebook fez modificações de modo que os anúncios do tipo criado por Julia Angwin não fossem mais possíveis. Depois de um artigo no *Guardian*, o Google aperfeiçoou seu preenchimento automático de modo que ele não apresentasse mais sugestões antissemitas, sexistas ou racistas. Apesar de o Google ter sido menos receptivo em relação ao trabalho de Amit Datta, ele conversou com a Microsoft sobre como ajudá-la a detectar discriminação em ofertas de trabalho on-line. O ativismo começou a fazer a diferença.

OS PRINCIPAIS ELEMENTOS DA AMIZADE

Talvez eu não seja um ativista típico. Sou um professor de matemática aplicada e faço parte da comunidade científica. Sou um inglês da classe média, de meia-idade, pai de dois filhos, que escapou da turbulência política de seu país natal para ter uma vida tranquila na Suécia. Contribuo para o desenvolvimento de algoritmos, e foi por isso que fui convidado para dar uma palestra no Google. Em meu trabalho, uso a matemática diariamente para entender melhor nossos hábitos sociais, para explicar como interagimos uns com os outros e descobrir as consequências de nossas interações. Mas não me envolvo muito com questões políticas.

Não tenho orgulho da minha passividade. Conversar com Angela Grammatas e outros como ela me fez perceber que eu estava mergulhado em meu laptop, ignorando os problemas. A ascensão dos algoritmos aconteceu numa época de incerteza crescente na Europa e nos Estados Unidos. Essas mudanças fazem as pessoas se sentirem dominadas. Quase toda reportagem — de Donald Trump ter usado a consultoria política Cambridge Analytica para selecionar eleitores durante sua campanha eleitoral à falha dos estatísticos ao preverem a votação do Brexit no Reino Unido — tem uma perspectiva algorítmica. Quando ouço meus amigos falarem sobre esses assuntos ou acompanho discussões no Twitter, percebo que não consigo responder adequadamente às questões que são levantadas.

As pessoas querem saber o que está acontecendo dentro das caixas-pretas que são usadas para nos avaliar e influenciar.

O termo “caixa-preta” foi usado tanto por Frank Pasquale, no título de seu livro *The Black Box Society*, quanto pela ProPublica, em sua série de matérias e vídeos curtos sobre algoritmos chamada “Breaking the Black Box”. É um conceito poderoso. Você insere seus dados, espera o modelo processá-los e recebe uma resposta. Você não consegue ver o que acontece lá dentro. A previsão de uma reincidência criminal é realizada por uma caixa-preta. As propagandas do Facebook e do Google são geradas por uma caixa-preta.

A ideia de caixa-preta pode nos dar um senso de impotência, um sentimento de que não somos capazes de entender o que os algoritmos fazem com nossos dados. Mas talvez esse sentimento seja ilusório. Podemos e devemos examinar o que acontece dentro dos algoritmos. Foi quando percebi que eu era capaz de fazer alguma coisa. Eu poderia analisar os algoritmos das caixas-pretas usadas em nossa sociedade e ver como eles funcionam. Mesmo não sendo um ativista, mas eu poderia responder algumas das perguntas que as pessoas estavam fazendo sobre as mudanças na sociedade.

Estava na hora de começar a trabalhar.

Pensei no que Angela Grammatas me disse sobre o Facebook: que era o site que mais sabia a nosso respeito. O gigante da mídia social era o melhor lugar para se começar a investigar como os algoritmos nos classificam. Eu precisava começar analisando algo que tinha certeza de que entendia completamente: minha própria vida social. Se eu criar uma caixa-preta modelo de meus amigos, eu deverei ser capaz de entender os passos realizados pelos cientistas de dados que trabalham no Facebook e no Google. Eu teria um contato direto com as técnicas que eles usam. A escala de meu modelo seria bem menor, mas a abordagem seria a mesma.

Angela estava certa. As páginas dos meus amigos no Facebook contêm uma imensa quantidade de informação sobre a vida deles. Eu abro o feed de notícias de meu Facebook e vejo uma atualização de um professor rabugento

reclamando que o condutor do trem freou abruptamente. Vejo fotos digitalizadas de uma festa que aconteceu há 25 anos numa escola. Vejo fotos de férias e de cervejas após o expediente. Vejo piadas sobre Donald Trump, campanhas para melhorar os sistemas de saúde e de habitação e indignações sobre decisões políticas. Vejo pessoas vangloriando-se do seu sucesso no trabalho e da criação dos filhos. Vejo fotos de casamento, retratos de bebês e crianças se divertindo em piscinas. Tudo pode ser encontrado no feed de notícias do Facebook, do extremamente pessoal ao explicitamente político.

Escolho 32 dos meus amigos no Facebook e analiso suas 15 publicações mais recentes.¹ Classifico cada publicação em uma das 13 categorias mais comuns: família/parceiro, atividades ao ar livre, trabalho, piadas/memes, produtos/propaganda, política/notícias, música/esporte/cinema, animais, amigos, eventos locais, pensamentos/reflexões, ativismo e estilo de vida. Depois construo uma matriz com 32 linhas e 13 colunas numa planilha, na qual eu preencho o número de vezes que meus amigos fizeram um tipo particular de publicação. Por exemplo, a linha correspondente a um dos meus amigos da universidade, Mark, tem uma publicação sobre seu trabalho, oito com fotos de férias em família, três fotos sobre a política do Brexit (sendo um escocês vivendo em Paris, ele é contra), uma publicação sobre uma viagem a Nova York e uma postagem marcando a si próprio como seguro após o ataque terrorista em Paris, em novembro de 2015. Na linha do meu colega Torbjörn, as publicações mais comuns (cinco delas) são sobre o jantar do Prêmio Nobel do qual ele não apenas participou como também foi entrevistado por uma rede de televisão sueca. Conto essas como postagens de trabalho, bem como outras duas publicações sobre palestras que ele ministrou. Torbjörn tem duas publicações de família e as demais estão espalhadas em várias outras categorias.

Para ter maior noção do equilíbrio entre o trabalho e a vida familiar de Mark, Torbjörn e outros amigos meus, fiz um gráfico do número de postagens deles sobre trabalho em função do número de publicações do tipo família/parceiro. O resultado é mostrado na Figura 3.1. Mark está na parte superior à esquerda, com oito publicações sobre família e uma sobre

trabalho. Torbjörn está na parte inferior à direita, com sete postagens sobre trabalho e duas sobre família. Cada um dos outros pontos representa como meus amigos estão posicionados nestas duas dimensões.

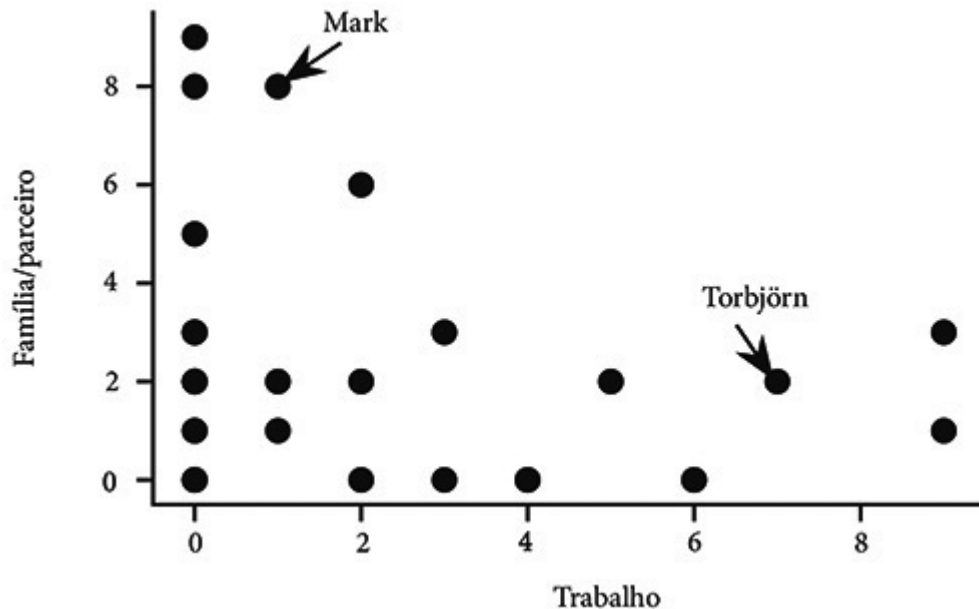


Figura 3.1 Classificando meus amigos nas dimensões de trabalho e família/parceiro. Cada ponto representa quantas vezes aquele amigo em particular publicou sobre cada assunto no Facebook.

Um pequeno grupo dos meus amigos se encaixou na categoria de publicar basicamente sobre trabalho, enquanto outros ficaram na categoria das postagens de família. Mas alguns publicaram sobre ambos os tópicos enquanto vários não abordaram muito nenhum dos dois. Cada categoria de publicação pode ser considerada uma dimensão do espaço, sendo que mostrei duas: a primeira dimensão é a de postagens sobre trabalho e a segunda é a sobre família. Mas eu poderia considerar uma terceira dimensão, a de publicações sobre atividades ao ar livre, uma quarta dimensão, a de política/notícias, e assim por diante. Cada um dos meus amigos corresponde a um único ponto deste espaço 13-dimensional.

O problema é que, ao aumentar o número de dimensões, fica cada vez mais difícil visualizar os dados. Não consigo formar uma ideia clara em

minha mente sobre como seria um ponto em um espaço 13-dimensional. Mas trabalhar com duas dimensões, como na Figura 3.1, não é o problema. Também é possível lidar com três dimensões: primeiro eu imagino os pontos colocados num cubo e, depois, penso em como os pontos mudariam de posição se eu girasse o cubo. Mas não conseguimos pensar em como o mundo se parece em quatro ou mais dimensões. Nosso cérebro é limitado a funcionar em duas ou três dimensões, uma vez que é dessa forma que vivenciamos nosso cotidiano.

Para nós, que não conseguimos ver pontos em quatro ou mais dimensões, a maneira mais simples de prosseguir é produzindo várias imagens bidimensionais instantâneas. A Figura 3.1 é um exemplo de imagem instantânea da relação entre trabalho e família/parceiro. De outras imagens instantâneas como esta, pude perceber que, para as pessoas que fazem postagens sobre estilo de vida, como alimentação e viagens, é incomum publicar também sobre política/notícias. Esses dois interesses estão correlacionados negativamente: amigos que compartilharam fotos de um restaurante recém-visitado tenderam a não dar opinião sobre temas atuais. Outros tipos de publicações são correlacionados positivamente: meus amigos que escreveram sobre música, cinema e esporte também tenderam a compartilhar piadas e memes.

A comparação de dados em pares começa a nos dar uma noção dos padrões existentes em um conjunto de dados 13-dimensional, mas esta não é uma abordagem particularmente sistemática. São 78 pares de relações para analisar,² então, representá-los graficamente e analisá-los é um processo demorado. Em alguns casos, há relações múltiplas: aqueles que compartilharam piadas e memes e escreveram sobre música e cinema também compartilharam notícias e política, mas tenderam a não fazer publicações pessoais sobre estilo de vida. Eu gostaria de ter uma maneira de classificar sistematicamente a intensidade dessas relações: descobrir quais são as mais importantes e quais capturam melhor as diferenças entre meus amigos.

Apliquei aos dados dos meus amigos um método conhecido por análise de componentes principais (ACP). O ACP é um método estatístico que rotaciona meu conjunto original de dados 13-dimensional, em que cada categoria de publicação é uma única dimensão, para revelar as relações mais importantes entre as publicações. A primeira componente principal, a relação que faz a correlação mais forte entre os dados, é uma reta que passa, no sentido positivo, pelas dimensões família/parceiro, estilo de vida e amigos, enquanto passa, no sentido negativo, pelas dimensões piadas/memes, política/notícias e trabalho. Essa é a relação mais importante que distingue meus amigos. Alguns gostam de publicar sobre o que têm feito no âmbito pessoal e outros gostam de compartilhar o que acontece no mundo e no trabalho.

A segunda relação mais importante entre os dados distingue trabalho de passatempos e interesses: passando, no sentido positivo, por trabalho e estilo de vida e, no sentido negativo, por música/esporte/cinema, política/notícias e outras publicações sobre cultura. Matematicamente, a segunda componente principal é a reta que está mais próxima dos pontos e que é perpendicular à primeira componente principal. Em 13 dimensões, é difícil visualizar o traçado de retas e a rotação de dados, mas determinar as retas e realizar as rotações são tarefas bastante simples para um computador.³ A Figura 3.2 mostra como 11 dos 13 tipos de publicações diferentes podem ser vistos num espaço bidimensional.



Figura 3.2 A primeira e a segunda componentes principais de uma análise das publicações dos meus amigos. Da esquerda para a direita, tem-se a primeira componente principal, que rotulei como “público versus pessoal”. De cima para baixo, tem-se a segunda componente principal, que rotulei como “cultura versus local de trabalho”. O tamanho da contribuição (negativo ou positivo) é indicado pelo comprimento do segmento de reta marcado na componente. Assim, por exemplo, “família/parceiro” é a publicação mais proeminente da primeira componente principal.

A maior contribuição positiva para a primeira componente principal (mostrada à direita na Figura 3.2) vem das publicações sobre família/parceiro, a segunda maior é sobre estilo de vida e as terceira e quarta são sobre amigos e atividades ao ar livre. O que essas publicações têm em comum é que são relacionadas à vida pessoal — elas exigem que escrevamos sobre as coisas que fazemos e as pessoas com quem as fazemos. As categorias piadas/memes, trabalho, música, esporte, cinema e política/notícias dão uma contribuição negativa a essa mesma componente principal (à esquerda na Figura 3.2). Publicações desse tipo estão relacionadas à vida pública: todas referem-se a trabalho ou a comentários sobre notícias ou atualidades. Chamei essa primeira componente principal de “público *versus* pessoal” porque ela captura as diferenças de como meus

amigos usam o Facebook tanto para publicar sobre eles mesmos como para comentar o mundo ao seu redor.

A maior categoria positiva que contribui para a segunda componente principal é o trabalho, seguido de estilo de vida (veja a parte superior da Figura 3.2). Muitas das publicações sobre estilo de vida que vejo no Facebook são sobre viagens de trabalho dos meus amigos — relaxando com uma cerveja depois de uma reunião ou fotos do jantar de uma conferência. Desse modo, faz sentido juntar essas duas categorias. Todas as pontuações negativas dessa categoria referem-se a eventos de uma esfera cultural mais ampla — notícias, esportes e piadas estão presentes, bem como ativismo e propaganda. Assim, essa segunda componente principal pode ser mais bem descrita como “cultura *versus* local de trabalho”.

Note que, embora eu atribua às componentes os nomes público *versus* pessoal e cultura *versus* local de trabalho, estou simplesmente nomeando as categorias geradas pelo algoritmo. Foi o algoritmo, não eu, que decidiu que essas eram as melhores dimensões para descrever meus amigos.

Agora que essas dimensões estão definidas, posso categorizar meus amigos. Quais estão mais interessados em vida pública e quais estão mais interessados em sua vida pessoal? Quais são os mais voltados para o trabalho e quais são os mais voltados para uma vida cultural?

Para descobrir, coloquei meus amigos em um gráfico bidimensional em que cultura *versus* local de trabalho está em função de público *versus* pessoal (Figura 3.3). Quando vi esses nomes aparecerem na minha tela, soube imediatamente que as componentes fizeram sentido. A maior parte das pessoas representadas por quadrados à direita — como Jessica, Mark e Ross — têm filhos e gostam de compartilhar informações sobre eles. A maioria das pessoas na parte inferior à esquerda, representadas por cruzes, não tinham filhos na época que fiz minha análise e escrevem mais sobre atualidades: Alma sobre literatura e teatro, Konrad sobre jogos para computador e Richard sobre política. O grupo superior à esquerda, representado por círculos, é tipicamente de acadêmicos que escrevem sobre seus trabalhos e artigos publicados recentemente. Torbjörn é um colega

biólogo matemático. Falarei mais tarde de Olle, um matemático ligeiramente excêntrico de Gotemburgo, que associa o interesse em política com trabalho em suas publicações.

O que mais me surpreendeu foi a profundidade com que essa classificação capturou similaridades e diferenças genuínas entre meus amigos. Lembre-se de que não informei ao algoritmo ACP como eu queria categorizar as pessoas. Forneci um conjunto amplo de 13 categorias que o ACP reduziu às duas dimensões mais pertinentes: público *versus* pessoal e cultura *versus* local de trabalho. E essas dimensões fazem sentido — as diferenças mais importantes entre meus amigos realmente se enquadram nessas dimensões.

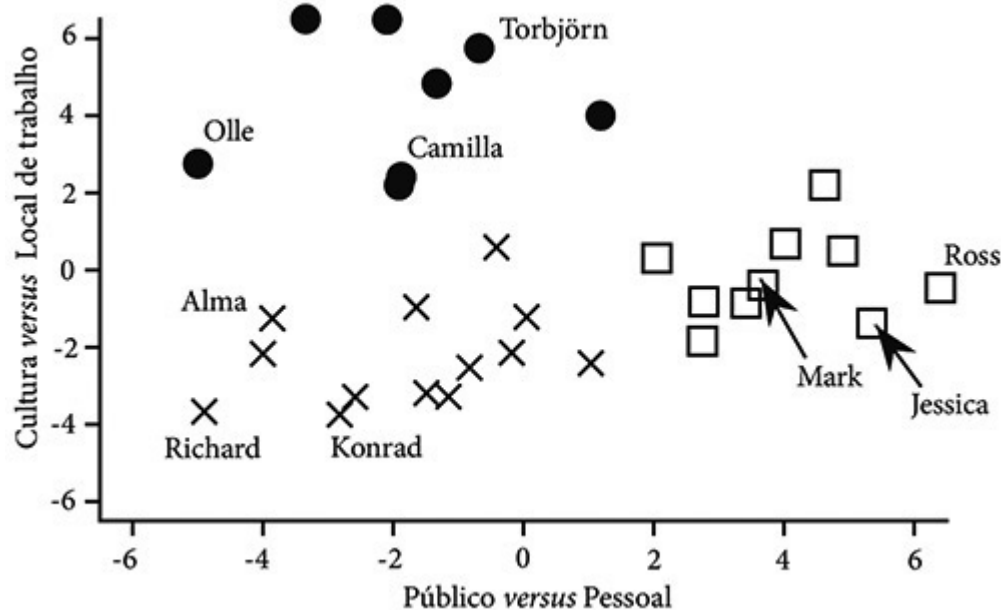


Figura 3.3 O detalhamento de meus amigos ao longo de duas componentes principais. As pessoas à direita (quadrados) publicam principalmente sobre amigos, família e vida pessoal. As pessoas na parte inferior à esquerda (cruzes) concentram suas publicações em notícias, esportes e vida pública. As pessoas na parte superior à esquerda (círculos) publicam principalmente sobre trabalho.

O agrupamento de meus amigos em três tipos distintos (círculos, quadrados e cruzes) também foi realizado por um algoritmo. Usei uma técnica computacional chamada “k-média” para agrupar indivíduos com base na

distância entre eles ao longo das dimensões criadas pelo ACP. Esse processo gerou três categorias: pessoas que usam o Facebook com foco em sua vida pessoal (quadrados); pessoas que focam em seu trabalho e estilo de vida relacionado ao trabalho (círculos) e pessoas que usam o Facebook para comentar sobre eventos variados da sociedade (cruzes). De novo, configurei meu algoritmo para encontrar a forma mais eficiente de agrupar meus amigos, e essas foram as categorias que ele gerou. A análise de componente principal usa os dados para classificar as pessoas, em vez de se basear em nossas preconcepções.

Os amigos que classifiquei concordaram em grande parte com as conclusões da minha análise de componentes principais. Camilla, que ficou na categoria trabalho, disse que a análise refletiu o modo como ela usa o Facebook, que é essencialmente para informar sobre sua vida profissional. Ela utiliza outras redes sociais para amigos e família. Já Ross usa a rede de maneira oposta: “somente para algumas fotos de família, como seu gráfico concluiu”, disse-me ele.

Torbjörn não gostou de ter sido classificado como “somente trabalho, nenhuma diversão”, mas concordou que, no Facebook, ele se concentrava na vida profissional e não na pessoal.

Classificar meus amigos do Facebook foi divertido, mas há um aspecto mais sério nesta redução de amizades a duas dimensões. A análise de componentes principais e abordagens matemáticas similares permeiam a maioria dos algoritmos utilizados para classificar nossos comportamentos. Esta abordagem é empregada em modelos sobre reincidência criminosa para transformar respostas dadas por um réu a um questionário em previsões sobre se ele ou ela cometerá mais crimes. É utilizada pelo Twitter para determinar quanto você ganha e é utilizada pelo Google para criar preferências de publicidade. A quantidade de dados envolvida e as dimensões nas quais somos classificados são muito maiores, mas a abordagem é a mesma que utilizei: gire e reduza até que o algoritmo comece a compreendê-lo.

É impressionante como apenas 15 publicações conseguem capturar nossas vidas. Imagine o que o Facebook consegue fazer com bilhões de publicações...

SUAS 100 DIMENSÕES

O Facebook tem dois bilhões de usuários no mundo inteiro, que produzem dezenas de milhões de publicações por hora, gerando um registro amplo de nossa atividade social. O professor Michal Kosinski, da Escola de Graduação em Negócios de Stanford, foi um dos primeiros pesquisadores a perceber que poderíamos usar a abordagem do ACP para categorizar pessoas com base na quantidade enorme de dados que elas estavam inserindo nas redes sociais. Enquanto cursava o doutorado na Universidade de Cambridge, desenvolveu, juntamente com David Stillwell, o projeto myPersonality [minha personalidade]. Eles coletaram um conjunto de dados impressionante: mais de três milhões de pessoas permitiram que eles acessassem e armazenassem seus perfis do Facebook. Várias dessas pessoas fizeram, então, uma série de testes psicométricos, avaliando sua inteligência, personalidade e felicidade, e responderam perguntas sobre sua orientação sexual, uso de drogas e outros aspectos de seu estilo de vida. As informações forneceram a Michal um enorme banco de dados que conecta como as coisas que escrevemos, compartilhamos e curtimos no Facebook se relacionam com nosso comportamento, nossas opiniões e nossa personalidade.

Michal começou analisando atributos em que podemos ser categorizados em uma de duas maneiras: republicano ou democrata, gay ou heterossexual,

cristão ou muçulmano, masculino ou feminino, solteiro ou em um relacionamento, e assim por diante. O objetivo de sua pesquisa era determinar se nossas “curtidas” podem ser usadas para avaliar quem somos: quais “curtidas” são mais prováveis de serem associadas a cada categoria?

A tabela que fornece algumas das “curtidas” previsíveis no artigo científico de Michal e seus colegas é uma lista de estereótipos terrivelmente vergonhosos.¹ Em 2010/11, quando o estudo foi realizado, homens gays curtiam Sue Sylvester da série de TV *Glee*, Adam Lambert do *American Idol* e apoiavam várias campanhas de direitos humanos. Homens heterossexuais curtiam Foot Locker, Wu-Tang Clan, X Games e Bruce Lee. Pessoas com poucos amigos curtiam o jogo *Minecraft*, hard rock e “passear com seu amigo e empurrá-lo para cima de alguém aleatoriamente”. Pessoas com muitos amigos curtiam Jennifer Lopez. Pessoas com QI baixo curtiam o personagem Clark Griswold, do filme *Férias frustradas*, “ser mãe” e as motocicletas Harley-Davidson. Pessoas com QI alto curtiam Mozart, ciência, *O senhor dos anéis* e *O poderoso chefão*. Os afro-americanos curtiam Hello Kitty, Barack Obama e a rapper Nicki Minaj, mas se mostravam menos interessados em acampar ou no político Mitt Romney do que outros grupos étnicos.

Essas observações não significam que devemos concluir que uma pessoa é gay com base numa única “curtida” para Sue Sylvester, ou que uma pessoa é inteligente só porque gosta de Mozart. Seria uma lógica bastante infantil: “Ha ha, você gosta de *Minecraft*... Você não tem nenhum amigo.” Tal raciocínio, além de desagradável, normalmente está errado.

Em vez disso, Michal descobriu que cada “curtida” fornece um pouco de informação sobre uma pessoa e que um acúmulo de “curtidas” permite que seu algoritmo faça conclusões confiáveis. Para agrupar todas as nossas “curtidas”, Michal e seus colegas usaram a análise de componentes principais. Ele reuniu todas as dezenas de milhões de diferentes “curtidas” das pessoas e usou a ACP para descobrir quais delas contribuíam para a mesma componente. Por exemplo, os Beatles, Red Hot Chili Peppers e o seriado de TV *House* ficaram todos juntos numa dimensão, que podemos

nomear de “rock e filmes de meia-idade”. Outra dimensão, que podemos chamar de “produtos anunciados”, inclui a Disney Pixar, Oreo e YouTube. E assim por diante. Michal descobriu que seriam necessárias de 40 a 100 dimensões para nos classificar com precisão.²

Michal enfatiza que os computadores descobrem mais relações sutis do que os humanos. “É fácil concluir que alguém que frequenta boates gays e compra revistas gays provavelmente é gay. Mas computadores podem tirar essas conclusões de sinais não tão evidentes para nós”, disse-me ele. De fato, somente 5% dos usuários rotulados como gays por esse algoritmo curtiram uma página explicitamente gay no Facebook. Foi a combinação de várias curtidas diferentes, variando de Britney Spears a *Desperate Housewives*, que permitiu ao algoritmo determinar a orientação sexual do usuário.

Embora a análise em larga escala de Michal, a partir de dados do Facebook, seja nova, o uso que ele faz da análise de componentes principais não é. Ao longo dos últimos 50 anos, sociólogos e psicólogos usaram a ACP para categorizar personalidades, valores sociais, opiniões políticas e status socioeconômicos. Gostamos de nos imaginar como sendo multidimensionais. Nós nos vemos como indivíduos complicados, com muitos aspectos diferentes de personalidade. Dizemos a nós mesmos que somos únicos, moldados pelos milhões de eventos específicos que ocorrem durante nossa vida. Mas a ACP pode reduzir esses milhões de dimensões de complexidade a um número muito menor de dimensões que podem ser usadas para nos colocar em caixas ou, usando uma metáfora visual mais apropriada, para nos representar com alguns símbolos diferentes. A ACP me diz que, de certa forma, posso enxergar meus amigos como aglomerados de círculos, quadrados ou cruzeiros.

Foi por meio dessa missão de nos enxergar como um aglomerado de símbolos que chegamos à lista de traços de personalidades que os psicólogos chamam de Big Five, ou Modelo dos Cinco Grandes Fatores.

A pesquisa dos psicólogos sobre personalidade baseia-se na nossa percepção diária sobre nossos amigos e conhecidos. Todos conhecemos pessoas que são amigáveis e comunicativas, que preferem estar entre outras

peessoas e que gostam de sair. Nós as chamamos de “extrovertidas”. Também conhecemos pessoas que gostam de ler e programar computadores, que apreciam ficar sozinhas e que falam menos quando estão em grupos. Nós as chamamos de “introvertidas”. Estas não são apenas noções frívolas. Elas são maneiras corretas e úteis de descrever pessoas.

Porém, muitas das nossas intuições sobre outras pessoas carecem de rigor científico. Existem muitas palavras que podemos usar para descrever nossos amigos e colegas: questionador, agradável, competente, submisso, idealista, determinado, disciplinado, depressivo, impulsivo... a lista continua. Confrontados com essa longa lista de adjetivos e depois de conduzir questionários nos quais os respondentes se avaliavam com base num grande número de afirmações diferentes — como, por exemplo, “Faço minhas tarefas logo” e “Falo com várias pessoas diferentes em festas” —, os psicólogos recorreram ao método ACP para descobrir os padrões fundamentais da nossa personalidade. Os resultados têm sido surpreendentemente consistentes. Percorrendo todas as dimensões dos adjetivos humanos e os tipos independentes de questões feitas, os psicólogos chegam, na maioria dos casos, sempre nas mesmas componentes da personalidade, as Big Five: abertura a novas experiências, conscienciosidade, extroversão, amabilidade (ou simpatia) e neuroticismo (ou instabilidade emocional).³

Esses cinco traços não são escolhas arbitrárias, e sim medidas robustas e reproduzíveis que resumem o que significa ser humano.

A questão seguinte ocorreu a Michal: se os traços de personalidade do Big Five eram sólidos e se as curtidas do Facebook podiam ser usadas para prever o QI e as opiniões políticas, também deveria ser possível prever personalidades a partir dos perfis do Facebook.

E era. Pessoas extrovertidas no Facebook gostam de dançar, de teatro e de jogar *Beer Pong*. Pessoas tímidas gostam de *anime*, de jogos de RPG e dos livros de Terry Pratchett. Pessoas neuróticas gostam de Kurt Cobain, de música emo e de dizer “algumas vezes me odeio”. Pessoas calmas gostam de paraquedismo, de futebol e de administração de empresas. Muitos

estereótipos foram confirmados pelas “curtidas” das pessoas, mas várias relações inesperadas também apareceram. Considero-me uma pessoa relativamente calma, inclinada a uma partida de futebol, mas dificilmente me jogaria de um avião, usando um paraquedas ou não. É o acúmulo de muitas curtidas diferentes que revela nossa personalidade, não apenas um único clique.

Estamos clicando nossa personalidade para dentro do Facebook, hora após hora, dia após dia. Carinhas felizes, dedão para cima, “curtir”, carinhas tristes e corações. Estamos dizendo ao Facebook quem somos e o que pensamos. Estamos nos revelando para uma rede social num nível de detalhes que normalmente reservamos para os amigos mais próximos. E, ao contrário de nossos amigos — que tendem a esquecer dos detalhes e são tolerantes em relação às conclusões que tiram sobre nós —, o Facebook está sistematicamente armazenando, processando e analisando nosso estado emocional. Ele está rotacionando nossa personalidade em centenas de dimensões, de modo que possa encontrar a direção mais fria e racional para nos analisar.

Os pesquisadores do Facebook dominaram as técnicas de reduzir nossa dimensionalidade. No estudo que fiz sobre meus próprios amigos, reduzi 13 dimensões de publicações de 32 pessoas para duas dimensões usando um algoritmo que demorou menos de um segundo para ser executado. Michal usou um algoritmo parecido, que levou cerca de uma hora para reduzir 55.000 curtidas diferentes de dezenas de milhares de pessoas para as cerca de 40 dimensões necessárias para prever a personalidade delas. O Facebook está trabalhando numa escala muito diferente. Seus métodos atuais levam menos de um segundo para reduzir um milhão de categorias distintas de curtidas, escolhidas dentre 100.000 pessoas diferentes, para poucas centenas de dimensões.⁴

Os métodos utilizados pelo Facebook são baseados na aleatoriedade. Rotacionar um milhão de vezes os dados de um milhão de categorias diferentes de curtidas — que foi o método que funcionou relativamente rápido para minhas 15 categorias — leva algum tempo. Por isso, o algoritmo

do Facebook começa escolhendo um conjunto aleatório de dimensões com as quais possa nos descrever. O algoritmo avalia, então, o desempenho dessas dimensões aleatórias, permitindo que ele encontre um novo conjunto de dimensões que melhore sua descrição. Depois de apenas algumas poucas iterações, o Facebook consegue ter uma ideia bastante boa das componentes mais importantes que descrevem seus usuários.

Se, por um lado, o Facebook consegue reduzir milhões de curtidas para umas poucas centenas de componentes, a visualização dessas componentes não é algo fácil para nós.⁵ Nosso cérebro, que funciona em duas ou três dimensões, e não em algumas centenas, atinge seu limite bem rapidamente. Assim, a fim de nos ajudar a entender como ele nos enxerga, o Facebook criou nomes para as categorias que seus algoritmos encontraram. Para descobrir como a empresa o caracterizou, primeiro você deve fazer o login; depois, clique no menu de opções na parte superior à direita e escolha “configurações”. Em configurações, escolha “anúncios”, então clique em “seus interesses” e, depois, clique em “mais”, última opção à direita; finalmente, escolha “estilo de vida e cultura”.

Quando o *New York Times* publicou um artigo explicando para as pessoas como encontrar essas preferências de anúncio, os leitores do jornal encontraram todos os tipos de categorias de interesse atribuídas a eles. Estão incluídas pessoas que foram categorizadas pelo Facebook em termos de seu interesse por “torrada”, “rebocadores”, “pescoço” e “ornitorrinco”.⁶

Dá para ver a graça disso, e as pessoas que tinham essas categorias deram boas risadas sobre como o Facebook as interpretou mal. Talvez isso seja verdade. Mas, quando vemos essas categorias, é importante lembrar que elas são uma tentativa de colocar palavras numa compreensão algorítmica muito mais profunda que o Facebook criou a partir de seus usuários. Os algoritmos que nos categorizam não são baseados em palavras. As palavras são alocadas para ajudar a nós, humanos, no entendimento das relações estatísticas entre os interesses das pessoas. De fato, essas relações não podem ser expressas em palavras como “torrada” e “ornitorrinco”. Elas não podem, de forma alguma, ser explicadas em palavras. Simplesmente não

conseguimos capturar a alta compreensão dimensional que o Facebook criou a partir de nós.

Em nossa conversa, Michal voltou a esse assunto repetidamente. Ele enfatizou que os humanos pensam sobre outras pessoas usando um número pequeno de dimensões — idade, raça, gênero e, se as conhecemos um pouco melhor, personalidade —, enquanto os algoritmos já estão processando bilhões de pontos de dados e realizando classificações em centenas de dimensões. Quando não entendemos como o Facebook nos classifica, a piada somos nós, não os algoritmos. Perdemos a capacidade de entender completamente as conclusões dos algoritmos que criamos.

“Somos melhores que os computadores em fazer coisas insignificantes que nós, por alguma razão, consideramos muito importantes, como andar por aí”, disse Michal. “Mas computadores podem realizar outras tarefas inteligentes que jamais poderemos fazer.” Do ponto de vista de Michal, a análise de componentes principais é o primeiro passo em direção à criação de um entendimento computadorizado e de alta dimensão da personalidade humana que supera a compreensão que temos atualmente sobre nós mesmos.

O Facebook obteve uma série de patentes que permitem à empresa utilizar sua compreensão multidimensional sobre nós. Uma das primeiras patentes foi para a formação de pares românticos.⁷ A ideia do Facebook é encontrar parceiros nos perfis dos usuários entre amigos de amigos. Volta e meia pensamos que nossos amigos solteiros, que talvez não se conheçam, podem formar um bom par. O sistema do Facebook poderia fazer essas sugestões para nós, baseado na personalidade estabelecida no perfil do usuário. A patente propõe permitir que usuários solteiros procurem por amigos de amigos para “identificar parceiros em potencial que satisfaçam a um conjunto de preferências de características desejadas, interesses ou experiências”. O amigo em comum pode ser, então, indagado se ele deseja ser um intermediário.

Se o Facebook pode achar um parceiro para você, então com certeza ele deveria ser capaz de encontrar um emprego para você. Em 2012, o

pesquisador Donald Kluemper e seus amigos descobriram que uma avaliação humana dos perfis do Facebook de 586 estudantes (principalmente mulheres brancas) forneceu avaliações confiáveis de seus prováveis locais de empregabilidade.⁸ Várias empresas terceirizadas solicitaram patentes para usar o Facebook e outras redes sociais para estender essa descoberta aos serviços automáticos de oferta e procura de emprego.⁹ A vantagem do uso do Facebook pelos empregadores, em relação a serviços puramente profissionais como o LinkedIn, é que o perfil do Facebook tem (para o bem ou para o mal) uma maior probabilidade de revelar seu verdadeiro eu.

O Facebook também está procurando maneiras de medir seu estado de espírito a partir de suas publicações, suas emoções a partir de suas expressões faciais em fotografias e seu nível de envolvimento a partir da sua taxa de interação com a tela.¹⁰ Pesquisas acadêmicas confirmaram que essas técnicas podem fornecer algum *insight* sobre seu estado de espírito. Por exemplo, a velocidade com a qual os usuários movem seu mouse durante uma tarefa rotineira no computador pode revelar o conteúdo emocional do que eles estão vendo na tela.¹¹ A análise de componentes principais pode detalhar o modo como você está interagindo com seu telefone ou computador a fim de construir uma imagem de como você está se sentindo.¹²

Esses avanços sugerem um futuro em que o Facebook monitora cada uma de nossas emoções e continuamente nos manipula em nossas escolhas como consumidores, nossos relacionamentos e nossas oportunidades de trabalho.

Se você usa o Facebook, Instagram, Snapchat, Twitter ou qualquer outra rede social regularmente, então você está dominado pelos números. Você está permitindo que sua personalidade seja tratada como um ponto em centenas de dimensões, que suas emoções sejam enumeradas e que seu comportamento futuro seja modelado e previsto. Tudo isso é feito efetiva e automaticamente, de uma maneira que a maioria de nós dificilmente pode entender.

A HIPERBOLITIZAÇÃO DE CAMBRIDGE

Depois da eleição presidencial americana de 2016, uma empresa chamada Cambridge Analytica anunciou que sua campanha orientada por dados tinha sido fundamental para a vitória de Donald Trump. A página principal do site da empresa apresentava uma montagem de notícias da CNN, CBSN, Bloomberg e Sky News mostrando a história de como ela havia usado a publicidade on-line direcionada e a pesquisa eleitoral em microescala para influenciar eleitores. O vídeo terminou com uma citação do especialista em pesquisas de opinião pública Frank Luntz: “Não há mais especialistas exceto a Cambridge Analytica. Eles foram o time de Trump, aqueles que descobriram como vencer.”

A Cambridge Analytica (CA) deu um grande destaque ao modelo de personalidade Big Five em seu material promocional. A empresa alega ter coletado centenas de milhões de pontos de dados sobre uma grande quantidade de eleitores americanos. A CA alegou que poderia usar esses dados para fornecer um quadro de personalidade dos eleitores que iria além das variáveis demográficas tradicionais, como gênero, idade e salário. Michal Kosinski, que conduziu o estudo sobre as personalidades do Facebook no capítulo 4, deixou bem claro, em nossa conversa, que ele não teve envolvimento algum com a CA, mas admitiu que a empresa poderia, sim, fazer uso de métodos similares aos que foram utilizados em sua pesquisa

científica a fim de influenciar eleitores. Se a CA tivesse acesso aos perfis do Facebook dos eleitores, ela poderia determinar quais tipos de propaganda teriam maior efeito sobre eles.

É um pensamento assustador. Os dados do Facebook podem ser usados para revelar nossas preferências, nosso QI e nossa personalidade. Essas dimensões poderiam então, pelo menos em teoria, ser utilizadas para nos direcionar mensagens que nos atraem como indivíduos: uma pessoa com o QI baixo pode ter sido bombardeada com teorias da conspiração não confirmadas sobre as contas de e-mail de Hilary Clinton; uma pessoa com o QI alto pode ter sido informada de que Donald Trump é um homem de negócios pragmático; uma pessoa com “afinidades afro-americanas” (como o Facebook as denomina) pode ter sido informada sobre uma restauração na periferia; um trabalhador branco desempregado pode ter recebido a informação de que um muro seria construído para manter imigrantes do lado de fora; e eleitores “com afinidade hispânica” podem ter sido informados sobre uma atitude firme em relação à Cuba de Castro. Personalidades neuróticas podem ser influenciadas com medo, pessoas solidárias, com empatia, e os extrovertidos, com um modo divertido de compartilhar a mensagem.

Em vez de focar numa mensagem central na mídia tradicional, o candidato de tal campanha pode se concentrar em desacreditar jornalistas e agências de notícias que estão tentando formar uma visão global da eleição. Com os meios de comunicação questionados, mensagens sob medida seriam enviadas diretamente aos indivíduos, fornecendo a eles uma propaganda que se adapta aos seus pontos de vista já estabelecidos.

Na época em que comecei a analisar a Cambridge Analytica, nos últimos meses de 2017, a empresa era muito mais cautelosa em como descrevia seu papel na vitória de Trump. Os jornais *Guardian* e *Observer* haviam investigado vários aspectos de como a CA coletou e compartilhou dados, ambos em conexão com a eleição presidencial dos Estados Unidos e o referendo sobre o Reino Unido deixar a União Europeia.¹ Como resultado, a CA estava, agora, se esforçando para minimizar o uso da psicologia em suas

campanhas. Em seu perfil, a companhia dizia que usava inteligência artificial para fazer uma segmentação de audiência, e não usava mais o termo “personalidade”.

Contatei o escritório de relações públicas da CA várias vezes para perguntar se eu poderia falar com algum membro da equipe técnica sobre como os algoritmos funcionavam. As respostas eram educadas, mas, por alguma razão, a pessoa com quem eu deveria falar estava sempre “ausente por causa do feriado prolongado” ou “de férias no momento”. Após uma enorme lista de desculpas, meus pedidos por e-mail não foram mais respondidos.

Então, decidi descobrir, por conta própria, como uma abordagem para vencer eleições baseada em personalidades políticas poderia funcionar.

Porém, antes de nos deixarmos levar pela ideia de políticos da direita se aproveitando de uma representação 100-dimensional dos eleitores americanos, precisamos refletir sobre quão acuradamente as dimensões dentro de um computador nos representam como pessoas.

Se eu quisesse insultar, de maneira infantil, a capacidade de pensar de um computador, eu poderia citar o fato de os computadores operarem em binário: em termos de zeros e uns. Mas esta é a forma errada de descrever modelos matemáticos. Na verdade, são os humanos que, frequentemente, enxergam as coisas em estados binários de preto e branco. Fazemos afirmações como “ele é burro demais para entender”, “ela é uma republicana típica” ou “esta pessoa compartilha tudo no Twitter” quase que reflexivamente, sem pensarmos na falta de sutileza de nossos julgamentos. São os humanos que veem o mundo de forma binária.²

Algoritmos bem projetados raramente categorizam eventos como pertencentes a uma categoria, entre duas. Eles trabalham, em vez disso, com classificações ou probabilidades. O modelo de personalidade do Facebook atribui uma classificação extrovertido/introvertido para cada usuário ou dá a probabilidade de um usuário ser “solteiro” ou estar “em um relacionamento”. Esses modelos lançam mão de uma gama de fatores e produzem um único

número, que é proporcional à probabilidade de um fato em particular sobre a pessoa ser verdadeiro.

O método mais básico utilizado para a conversão de grandes números de dimensões em uma probabilidade, ou classificação, é conhecido como regressão. Os estatísticos têm usado modelos de regressão por mais de um século, com aplicações começando na biologia e se estendendo até a economia, mercado de seguros, ciências políticas e sociologia. Um modelo de regressão pega os dados que já possuímos sobre uma pessoa e os usa para prever algo que não sabemos sobre ele ou ela. Para isso — um processo conhecido como ajuste do modelo —, precisamos primeiramente ter um grupo de pessoas sobre as quais já sabemos aquilo que estamos tentando prever.

Considere a relação entre idade e votação do Brexit. Dez dias antes da votação sobre a possibilidade de o Reino Unido deixar a União Europeia, a YouGov realizou uma pesquisa na qual perguntou às pessoas como elas iriam votar. Havia quatro grupos separados por idades incluídos na pesquisa: 18–24, 25–49, 50–64 e 65+, e verificou-se que as respostas dadas pelos entrevistados dependiam da faixa etária. A Figura 5.1 mostra um modelo de regressão que ajustei de acordo com as intenções de voto. À medida que a idade aumenta, aumenta a probabilidade de uma pessoa votar pela saída da União Europeia.

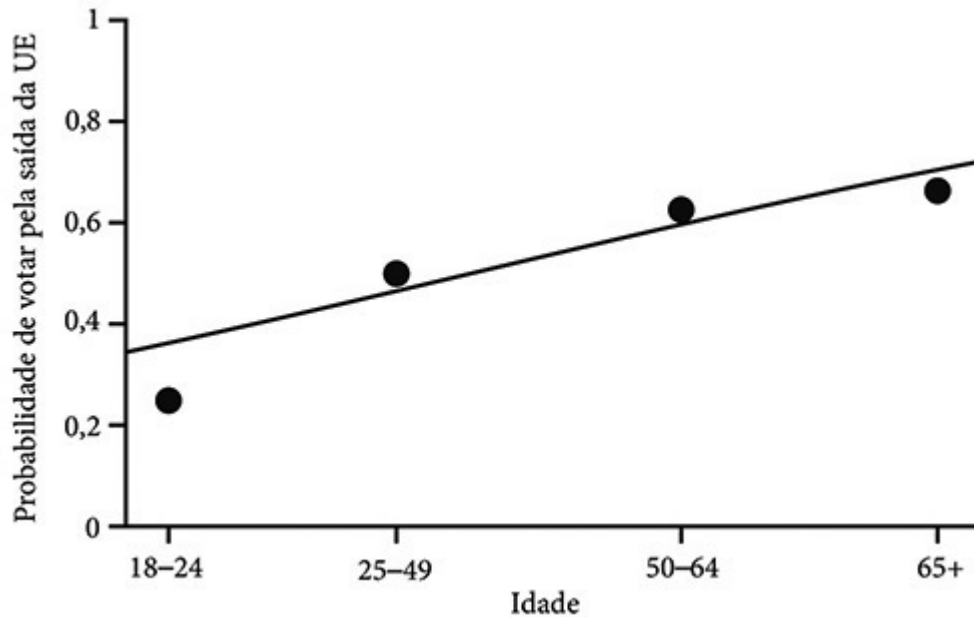


Figura 5.1 Modelo de regressão da probabilidade de uma pessoa votar pela saída da União Europeia em função da idade. Os círculos representam valores obtidos a partir dos dados da pesquisa de opinião coletada pela YouGov às vésperas da votação para decidir sobre a saída do Reino Unido da União Europeia em 2016.³ A linha sólida é um modelo ajustado, relacionando a idade à probabilidade de voto pela saída.⁴

O que as empresas de análise de dados usam para conseguir fazer previsões são modelos ajustados a um grupo de pessoas para inferir as preferências de outros. Dada a idade de uma pessoa, a Figura 5.1 pode ser usada para avaliar a probabilidade de ela votar pela saída da União Europeia. Baseado no modelo, essas empresas iriam inferir que a um jovem típico de 22 anos estaria associada uma probabilidade de 36% de voto pela saída, enquanto a um adulto típico de 60 anos estaria associada uma probabilidade de 62%.

Modelos de regressão não são uma representação perfeita dos dados. Na pesquisa, apenas 25% das pessoas com idades entre 18–24 disseram ser a favor do Brexit (veja os pontos na Figura 5.1). Portanto, o modelo superestima ligeiramente a probabilidade com que os jovens desejam a saída. Este tipo de inconsistência é típico de modelos de regressão que se propõem a representar um grande número de dados (neste caso, as idades das pessoas e as intenções de voto) em termos de uma única equação. Estou

levantando essa questão como uma observação, e não como um problema sério. A inconsistência não significa que o modelo esteja errado, apenas reflete uma limitação geral do método de regressão. Uma inconsistência pequena não é uma grande preocupação — todos os modelos estão errados até certo ponto e, no exemplo em questão, a quantidade de “erro” está dentro de níveis aceitáveis.

O único input de idade dá, ao meu modelo do Brexit, apenas um pequeno poder preditivo. Quanto mais inputs disponíveis, melhores as previsões serão. Para o voto sobre o Brexit, os analistas perceberam que pessoas mais velhas, pessoas com uma educação menos formal e pessoas da classe trabalhadora estavam mais predispostas a votar pela saída.⁵ Se uma agência contratada pela campanha “pró-saída” tivesse que escolher um grupo alvo de pessoas para encorajar a ir às ruas e votar, este seria o grupo de pessoas no qual ela deveria focar. Os defensores da permanência iriam preferir que os estudantes universitários fossem votar.

Os cientistas políticos têm usado a abordagem da regressão há muito tempo. Em um estudo conduzido após as eleições gerais de 1987 no Reino Unido, os pesquisadores mediram gênero, idade, classe social e percepção sobre a inflação de eleitores e analisaram como esses fatores afetavam as probabilidades de preferência pelo Partido Trabalhista ao Conservador.⁶ Os pesquisadores então mostraram que pessoas mais velhas e do sexo masculino eram muito mais propensas a votar nos Conservadores, enquanto pessoas da classe trabalhadora, com uma percepção de que a inflação estava alta, mostravam-se muito mais propensas a votar no Trabalhista. Os inputs para o modelo de regressão — gênero, idade, classe e percepção sobre a inflação — foram fornecidos ao modelo, e o resultado foi a probabilidade que uma pessoa votaria no Trabalhista.

A CA e outras empresas modernas de análise de dados usam mais ou menos as mesmas técnicas estatísticas usadas nos anos 1980. A principal diferença entre agora e antes está nos dados aos quais elas têm acesso. É possível fornecer aos modelos de regressão dados sobre as curtidas no Facebook, respostas a pesquisas on-line e dados sobre as compras que

fazemos. Em vez de se basear apenas na idade, classe e gênero para nos caracterizar, a CA diz utilizar estes grandes conjuntos de dados para estabelecer uma visão geral sobre nossa personalidade e posicionamento político. No passado, quando cientistas políticos estudavam as preferências partidárias dos eleitores, eles se baseavam, tipicamente, na formação socioeconômica. A CA diz: “Levar em conta o condicionamento comportamental de cada indivíduo [eleitor] para criar prognósticos fundamentados de comportamento futuro.”⁷

Para realizar uma regressão de larga escala na personalidade política das pessoas, a Cambridge Analytica precisou de muitos dados. Em 2014, o psicólogo Alex Kogan, um pesquisador da Universidade de Cambridge, estava coletando dados para seus estudos científicos por meio de um mercado de colaborações coletivas chamado Turco Mecânico. Alex descreveu o Turco Mecânico como “um grande agrupamento de pessoas que executam tarefas em troca de dinheiro”. Para seu estudo científico, ele pedia que as pessoas completassem um trabalho aparentemente insignificante: elas respondiam a duas perguntas sobre seu salário, informavam havia quanto tempo estavam no Facebook e, então, clicavam num botão para dar a Alex e seus colegas autorização para acessar os seus perfis no Facebook.

O estudo foi uma demonstração dramática do quão dispostas as pessoas estão a permitir que pesquisadores acessem seus próprios dados no Facebook, bem como os de seus amigos. Na ocasião, foi surpreendentemente fácil para os pesquisadores acessarem os dados da rede social. Ao obterem permissão dos colaboradores do Turco Mecânico, também foi possível acessar a localização e as curtidas dos seus amigos. Em troca de um dólar, 80% das pessoas que se voluntariaram para o estudo de Alex forneceram acesso a seus perfis, bem como os dados de localização de seus amigos. Os colaboradores tinham, em média, 353 amigos. Com apenas 857 participantes, Alex e seus colegas obtiveram acesso a um total de 287.739 dados pessoais. Este é o poder da rede social: dados coletados de um número reduzido de pessoas fornecem aos pesquisadores acesso aos dados de uma rede vasta de amigos.

Foi nesse momento que Alex começou a dialogar com representantes da SCL, um grupo de companhias que provê análises políticas e militares para clientes em todo o mundo. Inicialmente, a SCL estava interessada em ajudar Alex para a elaboração de questionários. Mas, quando os representantes da SCL perceberam o poder da coleta de dados no Turco Mecânico, a discussão voltou-se para a possibilidade de acessar uma enorme quantidade de dados de personalidades no Facebook. A SCL estava prestes a configurar o serviço de consultoria política, que posteriormente se tornaria a Cambridge Analytica, para usar as predições de personalidade a fim de ajudar seus clientes a vencerem eleições. Alex parecia possuir exatamente o acesso à coleta de dados de que a SCL precisava.

Alex admitiu ter sido ingênuo. Ele jamais havia trabalhado com uma empresa privada antes, tendo estado no ambiente acadêmico desde sua graduação em Berkeley, passando por seu Ph.D. em Hong Kong e chegando agora a sua posição de pesquisador em Cambridge. “Eu não fazia ideia de como os negócios são feitos”, ele me disse.

Ele e seus colegas consideraram os aspectos éticos e os riscos em trabalhar com a SCL, certificando-se de separar a coleta de dados de seu trabalho de pesquisa acadêmico. Eles perceberam que o Turco Mecânico não tinha a capacidade ou a segurança de coletar dados na escala necessária. Então eles usaram um já bem estabelecido serviço de pesquisas on-line ao consumidor, chamado Qualtrics. Alex me disse que, assim como nos estudos anteriores, eles pediram permissão para utilizar os perfis do Facebook dos entrevistados e seguiram todas as regras de acesso existentes na ocasião.

O que Alex não considerou foram os sentimentos e as percepções das pessoas quando elas souberam da coleta de dados do Facebook. “É bem irônico, se você pensar a respeito”, disse. “Muito do que eu estudo *são* emoções e eu acho que, se tivéssemos imaginado que as pessoas poderiam se sentir estranhas ou achar asqueroso fazermos previsões sobre a personalidade delas, então teríamos tomado uma decisão diferente.”

O jornal *Guardian* descobriu que, com o financiamento da SCL, uma empresa criada por Alex coletou dados do Facebook e respostas de

questionários de 200.000 cidadãos americanos.⁸ E isso representa apenas o número de pessoas que eles entrevistaram diretamente. Como a forma com que a plataforma do Facebook operava na época permitia o acesso às “curtidas” dos amigos das pessoas que se voluntariaram para o estudo e que consentiram o acesso aos dados de seus amigos, a SCL tinha no total dados de mais de 30 milhões de pessoas. Esse era um conjunto de dados imenso que fazia, potencialmente, um retrato da personalidade política de muitos americanos.

O CEO da Cambridge Analytica, Alexander Nix, não pareceu particularmente preocupado com o fato de pessoas acharem “asqueroso” sua empresa prever a personalidade política delas quando ele apresentou a pesquisa no Summit de Concordia em 2016.⁹ A empresa havia acabado de ajudar o presidenciável Ted Cruz a sair da obscuridade para se tornar um forte candidato às primárias republicanas. Nix explicou como, em vez de selecionar pessoas com base em raça, gênero ou formação socioeconômica, sua companhia conseguia “prever a personalidade de cada um dos adultos nos Estados Unidos da América”. A eleitores bastante conscienciosos e neuróticos poderia ser direcionada a mensagem de que “a segunda emenda era uma apólice de seguros”. Eleitores tradicionais e cordatos seriam informados de como “o direito de portar armas era importante e deveria ser transmitido de pai para filho”. Ele afirmou que conseguia usar “centenas e milhares de pontos de dados individuais em nossas audiências selecionadas para compreender exatamente quais mensagens teriam afinidade com quais audiências” e sugeriu que os métodos que ele havia descrito estavam sendo utilizados na campanha de Trump.

A origem da Cambridge Analytica apresenta todos os ingredientes de uma história de conspiração moderna. Ela envolve Ted Cruz, Donald Trump, segurança de dados, psicologia da personalidade, o Facebook, trabalhadores mal pagos do Turco Mecânico, Big Data, acadêmicos da Universidade de Cambridge, o populista de direita Steve Bannon, que faz parte da diretoria, o financiador de direita Robert Mercer, que é um dos seus maiores investidores, o ex-conselheiro de segurança nacional Michael Flynn,

que já atuou como consultor, e (em uma versão menos confiável desta história) *trolls* financiados pela Rússia. Eu consigo imaginar essa história como um filme com Jesse Eisenberg no papel de um psicólogo que gradualmente desvenda a motivação real da empresa para a qual trabalha: manipular cada uma de nossas emoções para fins políticos.

Olhando por esta ótica, esta é uma história assustadora. Mas quando me concentrei nos detalhes dos modelos usados para prever padrões de voto, percebi que um ingrediente importante estava faltando: o algoritmo. Eu queria conferir se as afirmações de Nix sobreviveriam ao escrutínio.

Não tenho acesso aos dados coletados por Alex Kogan — retornarei ao que pode ter acontecido com eles mais tarde —, mas Michal Kosinski e seus colegas criaram um pacote tutorial que permite a estudantes de psicologia praticar a criação de modelos de regressão em um banco de dados anônimo de 20.000 usuários do Facebook. Eu baixei o pacote e o instalei no meu computador. Apenas 4.744 dos 19.742 usuários do Facebook residentes nos Estados Unidos no banco de dados expressou a preferência por democratas ou republicanos.¹⁰ Destes, 31% eram republicanos. Na ocasião da coleta de dados, entre 2007 e 2012, os democratas se destacavam no Facebook. Usei os dados para ajustar um modelo de regressão com as 50 dimensões do Facebook como input. O resultado do modelo de regressão é a probabilidade de que uma pessoa seja republicana.

Após ajustar o modelo aos dados, o próximo passo é testar sua performance. Uma boa maneira de testar a acurácia de um modelo de regressão é selecionar duas pessoas aleatoriamente, um democrata e um republicano, e pedir ao modelo que preveja qual dos dois é o republicano a partir do seu perfil do Facebook. Esta é uma medida intuitiva de acurácia. Imagine que você encontrasse essas duas pessoas e pudesse fazer-lhes algumas perguntas sobre seus gostos e hobbies, e, a partir das respostas, você tivesse que determinar qual pessoa apoiava qual partido político. Com que frequência você acha que acertaria?

A acurácia de um modelo de regressão baseado nos dados do Facebook é muito boa. Em oito de nove tentativas o modelo de regressão identificou

corretamente as visões políticas do usuário do Facebook. O principal grupo de curtidas que identifica um democrata inclui o casal Barak e Michelle Obama, a Rádio Pública Nacional, TED Talks, Harry Potter, a página da internet I Fucking Love Science e shows de variedades atuais liberais como *The Colbert Report* e *The Daily Show*. Os republicanos curtem George W. Bush, a bíblia, música country e acampar.

Não é nenhuma surpresa que os democratas curtam os Obama e *The Colbert Report* ou que muitos republicanos curtam George W. Bush e a bíblia. Então, eu tentei ver se conseguia quebrar o modelo de regressão retirando algumas das “curtidas” óbvias do modelo e executar uma nova regressão. Para meu espanto, o modelo continuou funcionando com 85% de acurácia, com apenas uma ligeira redução em performance. Agora ele utilizava combinações de curtidas para determinar as filiações políticas. Por exemplo, alguém que curta Lady Gaga, Starbucks e música country se encaixa mais provavelmente como republicano, mas um fã de Lady Gaga que também gosta de Alicia Keys e Harry Potter se encaixa mais provavelmente como um democrata. É aí que a compreensão multidimensional, obtida com a utilização de muitas “curtidas”, produz resultados inesperados e úteis.

Este tipo de informação poderia ser bastante útil para um partido político. Em vez de os democratas direcionarem sua campanha apenas para a média liberal tradicional, eles poderiam se dedicar a obter os votos dos fãs de Harry Potter. Os republicanos poderiam alvejar pessoas que bebem café na Starbucks e pessoas que acampam. Os fãs de Lady Gaga deveriam ser tratados com cautela por ambas as partes. Apesar de ser difícil fazer uma comparação direta, a acurácia de um modelo de regressão baseado no Facebook parece vencer os métodos tradicionais. Por exemplo, no estudo da eleição geral do Reino Unido em 1987, pesquisadores concluíram que a probabilidade de um homem de classe média com 65 anos de idade, que considerava a inflação baixa, preferir votar no partido Conservador em vez do Trabalhista era de 79%. Assim, um modelo que supusesse que os “tóris

típicos” seriam, de fato, simpatizantes dos conservadores estaria errado pelo menos 21% das vezes.

Até aqui, tudo bem para Alexander Nix e a Cambridge Analytica. Mas, antes de nos empolgar, vamos olhar com mais atenção para as limitações.

Em primeiro lugar, há uma limitação fundamental nos modelos de regressão. Lembre-se de que o resultado de algoritmos não é binário. Tampouco, como vimos na Figura 5.1, um modelo é uma representação perfeita dos dados. Não podemos esperar que um modelo revele suas visões políticas com 100% de certeza. Não há como a Cambridge Analytica, ou quem quer que seja, tirar conclusões com acurácia garantida apenas ao analisar seus dados do Facebook. A não ser que você seja o Barack Obama ou a Theresa May. Ao contrário, o melhor que os analistas conseguem fazer é usar um modelo de regressão que atribui uma probabilidade sobre você ter uma visão em particular.

Enquanto os modelos de regressão funcionam muito bem para democratas ou republicanos convictos — como estabeleci anteriormente, a acurácia é de cerca de 85% —, as previsões sobre estes eleitores não são particularmente úteis em uma campanha política. Os votos dos simpatizantes partidários são mais ou menos garantidos, e não é preciso tê-los como alvo. Na verdade, o modelo de regressão que ajustei com os dados do Facebook não revelam nada sobre os 76% das pessoas que não registraram sua fidelidade política. Se, por um lado, os dados nos mostram que os democratas tendem a gostar de Harry Potter, por outro lado, eles não necessariamente nos dizem que outros fãs de Harry Potter gostam dos democratas. Este é o problema clássico inerente a todas as análises estatísticas; de uma possível correlação confusa, sem causa.

Uma segunda limitação diz respeito ao número de “curtidas” necessárias para se fazer uma previsão. O modelo de regressão só funciona quando uma pessoa deu mais de 50 “curtidas” e, para que esta previsão seja realmente confiável, algumas centenas de “curtidas” são necessárias. No conjunto de dados do Facebook, apenas 18% dos usuários “curtiram” mais de 50 páginas. Após a coleta desses dados, o Facebook conseguiu aumentar o número de

páginas que seus usuários “curtem”, exatamente para que possa melhorar o direcionamento da propaganda. Mas ainda há muitas pessoas, inclusive eu, que não “curtem” muito no Facebook. Eu “curto” um total de quatro páginas: a minha própria página *Soccermatics*, uma reserva natural local, a escola do meu filho e European Union Research. Não importa quão boa seja uma técnica de regressão, um modelo não consegue funcionar sem dados.

A terceira limitação está no coração da ideia de Nix de ter como alvo nossa personalidade política: será que um algoritmo consegue distinguir de modo confiável uma pessoa neurótica ou solidária a partir de suas “curtidas”? O conjunto de dados que utilizei incluía os resultados de um teste de personalidade que media os cinco traços de personalidade do Big Five. Usei isso para verificar se um modelo de regressão conseguiria determinar qual indivíduo, escolhido de um par aleatório, era mais neurótico. Simplesmente não funcionou. Escolhi duas pessoas do conjunto de dados de maneira aleatória e olhei para as suas pontuações de neuroticismo, a partir do teste de personalidade que executaram. Comparei essas pontuações com um modelo de regressão baseado nas “curtidas” do Facebook. O teste de personalidade e o modelo de regressão produziram as mesmas classificações para estes pares em apenas 60% dos casos. Se eu tivesse configurado as pontuações ao acaso, teria acertado 50% das vezes. O modelo foi apenas ligeiramente melhor do que a aleatoriedade.

O modelo de regressão foi um pouco melhor ao classificar as pessoas em termos de sua abertura a experiências: acertou cerca de dois terços das vezes. Mas quando realizei o mesmo teste em relação à extroversão, conscienciosidade e amabilidade, obtive resultados similares ao neuroticismo: o modelo acertou seis vezes de dez, comparado com as cinco vezes de dez que esperaríamos se tivéssemos designado pessoas aleatoriamente.

Foi nesse momento que comecei a discutir os resultados com Alex Kogan, o psicólogo da Cambridge que a princípio ajudou a CA na coleta de dados. De início, ele relutou em falar comigo, uma vez que se sentiu retratado de maneira injusta no artigo do *Guardian* e em um número de

blogs on-line sobre a Cambridge Analytica.¹¹ Mas quando contei a ele a respeito de minhas descobertas sobre prever personalidade com o Facebook ele começou a se abrir.

Alex havia chegado a conclusões similares às minhas. Ele não acreditava que a Cambridge, ou quem quer que seja, pudesse produzir um algoritmo que classificasse eficientemente a personalidade das pessoas. Ele estava trabalhando com uma combinação de simulações de computadores e dados do Twitter para mostrar que, apesar de aspectos de personalidade poderem ser medidos a partir de nossas pegadas digitais, o sinal não era forte o suficiente para fazer previsões confiáveis sobre nós. Ele foi franco em relação a Alexander Nix. “Nix está tentando promover [o algoritmo de personalidade] porque ele tem um grande incentivo financeiro para contar uma história sobre como a Cambridge Analytica tem uma arma secreta.”

Há uma distinção importante a ser feita aqui entre uma descoberta científica — de que um certo conjunto de “curtidas” no Facebook está relacionado com o resultado de testes de personalidade — e a implementação de um algoritmo confiável baseado nessa descoberta, criando uma equação que preveja corretamente que tipo de pessoa você é. Uma descoberta científica pode ser verdadeira e interessante, mas, a menos que a relação seja muito forte (o que não é o caso das previsões de personalidade), ela não nos permite fazer previsões particularmente confiáveis sobre o comportamento de um indivíduo.

Uma razão para que a distinção entre resultados científicos e algoritmos aplicados fique confusa é a forma com que resultados como esses são transmitidos pela mídia. Em janeiro de 2015, a revista *Wired* publicou um artigo intitulado: “Como o Facebook o conhece melhor do que os seus próprios amigos.” O jornal *Telegraph* do Reino Unido foi um passo além e estampou a manchete: “O Facebook o conhece melhor do que os membros da sua própria família.” Para não ficar de fora, o *New York Times* superou todas as outras mídias ao publicar: “O Facebook o conhece melhor que todo mundo.”

Todas essas manchetes foram seguidas por uma reportagem sobre o mesmo artigo científico. A pesquisa — conduzida por Wu Youyou, Michal Kosinski e David Stillwell — analisou o quão bem as “curtidas” do Facebook previram respostas às questões sobre personalidade, mas, dessa vez, compararam um modelo de regressão linear baseados em curtidas com respostas de um questionário de dez itens que colegas de trabalho, amigos, parentes e parceiros preencheram sobre o usuário do Facebook. O resultado científico, que os jornais estavam tentando capturar em suas manchetes, foi que o modelo estatístico deles se correlacionava melhor com o teste de personalidade do que com as dez respostas dadas por amigos e familiares.

Uma correlação melhor implica uma predição melhor, mas isso significa que o Facebook o conhece melhor do que qualquer outra pessoa? Claro que não. Perguntei a Brian Connelly, professor associado do Departamento de Administração da Universidade de Toronto, que estuda personalidade no ambiente de trabalho, o que ele achava sobre a pesquisa. “O trabalho de Michal [Kosinski] é interessante e provocativo, mas eu acho que a mídia está agindo de maneira sensacionalista”, ele me disse. “Uma manchete mais apropriada como ‘Descobertas preliminares sugerem que o Facebook conhece alguns de vocês quase tão bem quanto uma pessoa próxima (mas estamos aguardando para ver se o Facebook pode prever seu comportamento)’ não é muito impactante.” A manchete reformulada de Brian resume tudo. A ciência é interessante, mas ainda não há evidências de que o Facebook pode determinar e acertar sua personalidade política.

A história da Cambridge Analytica me fez mergulhar fundo numa rede de blogs e sites de ativistas pró-privacidade. Seguindo estes links, deparei-me com um vídeo do YouTube de um jovem cientista de dados que agora trabalha para a Cambridge Analytica, no qual apresenta um projeto de pesquisa que desenvolveu quando era estagiário na empresa. Ele começa sua apresentação com uma referência ao filme *Ela*, no qual o protagonista Theodore, interpretado por Joaquin Phoenix, se apaixona por seu sistema operacional (OS). No filme, o computador tem um entendimento profundo da personalidade de Theodore, e o humano e o OS se apaixonam. O jovem

cientista de dados usa essa história para abrir sua apresentação de cinco minutos: “Um computador pode nos entender melhor que um humano?”

O cientista de dados diz que sim. A apresentação dele nos leva, passo a passo, pela pesquisa de atividades on-line e personalidade. Ele descreve os traços de personalidade Big Five; como pesquisas de opinião podem ser substituídas por perfis do Facebook; afirma que os resultados de um de seus modelos de regressão revelam nossos níveis de conscienciosidade e neuroticismo; ele fala como mensagens políticas podem ser direcionadas aos indivíduos e, então, ele finaliza afirmando que “meu modelo, a partir de suas curtidas do Facebook, sua idade e gênero, pode prever seu nível de amabilidade bem como o de seu parceiro”. Um dia, ele diz, podemos nos apaixonar por um computador que nos entende melhor que nosso parceiro.

Começo a duvidar se o cientista de dados do vídeo realmente acredita no que está dizendo. Não tenho certeza se ele sequer espera que sua plateia acredite nele. Sua “pesquisa” é o produto de um programa de treinamento de oito semanas na ASI Data Science, um programa para aspirantes a cientistas de dados. Mas, mesmo que tenha sido apenas uma apresentação descompromissada, fiquei profundamente perturbado com o que vi. Trata-se de um jovem com o maior nível de treinamento científico: um doutorado em física teórica na Universidade de Cambridge. Assim, acho difícil de acreditar que ele não tenha algumas das mesmas dúvidas que eu. “Você se baseia em quais dados?”, gostaria de perguntar a ele. “Você testou e validou seu modelo ao longo do tempo? O que me diz sobre o fato de curtidas medirem o neuroticismo apenas ligeiramente melhor do que o método aleatório?”

Parece que a bolsa de estudos da ASI o encorajou a colocar essas dúvidas de lado quando ele apresentou seu projeto de pesquisa. Como resultado pessoal, ele recebeu uma oferta de emprego da Cambridge Analytica, que aceitou prontamente.

Não conheço esse jovem, mas tenho contato com vários outros como ele. Trabalho com eles e os oriento como estudantes de doutorado, de mestrado e de graduação. O que senti, quando assisti ao vídeo, foi uma sensação de

fracasso profundo da minha parte. Existe uma demanda de empresas como a Cambridge Analytica, e as universidades fornecem a elas esse tipo de perfil jovem e ambicioso: pessoas que podem ao mesmo tempo fazer pesquisa e apresentar resultados de um modo fácil de ser entendidos.

Vivemos uma época excitante, em que usamos dados para nos ajudar a tomar as melhores decisões e manter as pessoas informadas sobre questões que são importantes para elas. Mas com esse poder vem a responsabilidade para explicar cuidadosamente o que podemos ou não fazer. Quando meus colegas e eu treinamos pesquisadores, falamos sobre seus poderes, mas com frequência esquecemos de falar sobre suas responsabilidades. Parece que deixamos essa tarefa importante nas mãos dos consultores do mercado que estão ensinando a cientistas de dados como interpretar sua pesquisa para obter o maior efeito possível.

Ou o jovem do vídeo está dominado pelos números — incapaz de enxergar as limitações do método que apresenta — ou ele está tentando dominar o público — ao deixar de mencionar suas limitações. Em vez de afirmar que “meu” algoritmo conhece você tão bem quanto seu parceiro, um cientista cuidadoso diria “para indivíduos que usam bastante o Facebook, um estudo realizado por Michal Kosinski e seus colegas mostrou que suas curtidas podem ser usadas para prever pontos de sua personalidade, mas o que isso significa sobre a propaganda baseada em personalidade ainda não está claro”. Infelizmente, como a manchete de Brian Connolly, esta última afirmação não tem o mesmo impacto. Nem é um assunto que seu futuro empregador quer ter como parte de uma demonstração de cinco minutos da metodologia deles. Ciência cuidadosa não vende um serviço de consultoria política.

Poucos meses após Trump assumir a presidência, a Cambridge Analytica removeu a referência ao modelo de personalidade Big Five de seu site. Soube por uma fonte confiável que o Facebook mandou eles deletarem todos os dados que coletaram sobre as “curtidas” de seus usuários *antes* de começarem a trabalhar na campanha de Trump. CA alega que cumpriu a ordem do Facebook. É pouco provável que eles sequer tenham tentado levar

adiante o direcionamento descrito por Alexander Nix no Summit de Concordia, em conexão com a campanha de Trump. Desde então, a CA afirma que nenhum dos dados do Facebook que ela recebeu de Alex Kogan foi usado nos serviços que forneceu à campanha de Trump, nem que propaganda direcionada à personalidade foi usada na campanha em geral.

Em janeiro de 2017, David Carroll, professor associado da Parsons School of Design, em Nova York, solicitou uma proteção de dados à Cambridge Analytica. A CA respondeu com uma lista de informações que tinha sobre ele. O arquivo continha sua idade, gênero e local de residência. Havia uma planilha mostrando o distrito no qual ele havia votado, incluindo uma coluna indicando que votara nas primárias democratas. Eles usaram esses dados para classificar a importância que achavam que ele daria a vários assuntos, como meio ambiente, saúde e dívida nacional. Eles perceberam que era “muito improvável que fosse republicano”, com propensão “muito alta” a votar nas eleições. Depois de toda a propaganda exagerada, a Cambridge Analytica estava usando métodos de regressão tradicionais baseados em idade e local de residência para prever o voto de David. Os dados obtidos e os métodos usados estavam bem longe de ser a propaganda política pessoal direcionada sobre a qual Alexander Nix tinha se vangloriado.

A história da Cambridge Analytica* é, na minha opinião, fundamentalmente sobre hipérbole. É uma história sobre uma empresa aparentemente exagerando o que consegue fazer com dados. Alexander Nix admitiu “falar um pouco hiperbolicamente” sobre o que a CA faz. Mas esse é apenas um estudo de caso. Com todos, do Facebook e Spotify a agências de viagens e consultores de esporte, parecendo oferecer algoritmos que nos classificam e explicam nossos comportamentos, eu precisava descobrir mais sobre a acurácia desses algoritmos. Quão bem esses algoritmos realmente nos conhecem? Eles estão cometendo outros erros, até mais perigosos?

Nota

* A hipérbole criada pela CA sobre sua abordagem explodiu em grande escala quando esse livro foi enviado para publicação no Reino Unido em 2018. Para mais informações, veja as notas finais.¹²

IMPOSSIVELMENTE IMPARCIAL

Dissecar algoritmos de personalidade mudou minha perspectiva, mas não da maneira que eu esperava. Eu estava menos preocupado com o fato de algoritmos fazerem previsões perigosamente acuradas sobre nós e mais preocupado em como eles estavam sendo apresentados. As conclusões às quais cheguei sobre a Cambridge Analytica eram parecidas com aquelas a que cheguei, de um modo mais hesitante, após ler o artigo de Banksy. Neste último caso, os pesquisadores precisaram identificar Banksy a fim de localizá-lo. Algoritmos são úteis na organização de dados em uma campanha política ou em uma investigação criminal, mas não era simplesmente o caso de se apertar um botão e encontrar um artista de rua ou uma lista de republicanos neuróticos.

Algoritmos são apresentados, corriqueiramente, como fornecedores de insight de como somos como pessoas e capazes de prever como nos comportaremos no futuro. Eles são usados para determinar se seremos selecionados ou não para um emprego, se conseguiremos um empréstimo ou se deveríamos ser presos. Percebi que eu precisava saber mais sobre o que acontece dentro desses algoritmos e os tipos de erros que eles podem cometer.

Nos Estados Unidos, o algoritmo Compas é usado em alguns estados para ajudar a realizar as avaliações de risco de acusados, geralmente quando

estão tentando conseguir liberdade condicional. Algumas reportagens da mídia descreveram o Compas como uma caixa-preta, insinuando que é difícil ou impossível saber o que acontece lá dentro. Entrei em contato com Tim Brennan, criador do algoritmo Compas e diretor da Northpointe, empresa que o disponibiliza, e lhe perguntei se estaria disposto a explicar como o modelo funciona. Depois da troca de alguns e-mails, ele me enviou os relatórios internos que explicam como as pontuações foram criadas.¹ Mais tarde, quando o entrevistei, ele foi razoavelmente transparente sobre o modelo, me mostrando as equações específicas necessárias para entendê-lo.

O modelo de Tim usa uma combinação do histórico do acusado, idade na primeira prisão, idade atual e nível de educação, com as respostas de um questionário de uma hora de duração realizado durante uma entrevista para prever se ele vai reincidir. Todas essas medidas são usadas para ajustar um modelo estatístico baseado em antecedentes criminais. Pessoas com um histórico de desobediência ou violência têm maior probabilidade de reincidência, assim como pessoas com nível de educação baixo ou que usam drogas.² Indivíduos com problemas financeiros ou que se mudam muito *não* têm uma probabilidade maior de reincidência. São esses padrões da população como um todo que o modelo usa para fazer previsões.

As técnicas usadas dentro do Compas são parecidas com aquelas que tenho usado até então: primeiro, os dados são rotacionados e reduzidos usando o ACP e, então, um modelo de regressão é usado para prever reincidência com base em resultados históricos. Eu não diria que foi fácil acompanhar os detalhes sendo alguém de fora. Os relatórios técnicos tinham centenas de páginas, mas o modelo estava totalmente documentado, e Tim me chamou a atenção para as partes mais relevantes. Depois de ter lidado com a Cambridge Analytica, eu estava impressionado com a transparência da Northpointe.

O fato de os criadores de um algoritmo serem transparentes sobre os detalhes não significa que o algoritmo deles obtenha resultados corretos. Em 2015, um artigo da Julia Angwin alegou que o algoritmo era preconceituoso em relação a afro-americanos.³ A ProPublica usou o único método infalível

para determinar se um algoritmo é ou não imparcial: examinar a qualidade das previsões que ele faz. O Compas atribui uma pontuação de um a dez relacionada à probabilidade de o acusado ser preso por um delito no futuro. Os resultados de Julia e seus colegas foram claros. Eles descobriram que 45% dos acusados negros que receberam uma pontuação de risco mais elevado e, portanto, com maior probabilidade de serem presos, também tinham sido colocados numa categoria de alto risco. Isso comparado a uma faixa de erro de 23% para criminosos brancos. Acusados negros que não cometeram outros delitos tinham uma chance maior de serem classificados como alto risco do que os acusados brancos.

Quando Julia e seus colegas publicaram o artigo, Tim e a Northpointe responderam rapidamente. Eles escreveram um relatório de pesquisa afirmando que a análise da ProPublica estava errada.⁴ Argumentaram que o Compas obedece às mesmas normas que outros algoritmos examinados e testados. Alegaram que Julia e seus colegas interpretaram mal o que significa um algoritmo cometer um erro e que o algoritmo deles era “bem calibrado” para acusados brancos e negros.

O debate entre a Northpointe e a ProPublica me fez perceber quão complicado é o tema da parcialidade. Essas pessoas são inteligentes, e suas trocas de palavras preencheram quase 100 páginas de argumentos e contra-argumentos. Elas foram complementadas com um código de computador e análises estatísticas adicionais. O debate dos dois interlocutores foi, então, comentado por blogueiros, matemáticos e jornalistas, e cada um deles adicionou sua própria perspectiva ao problema. Definir parcialidade era um problema matemático difícil, o qual eu precisava examinar em detalhes para poder entendê-lo.

Assim, eu baixei os dados que a ProPublica tinha coletado e fui trabalhar.

Para entender o argumento da ProPublica e o contra-argumento da Northpointe, comecei refazendo a tabela que mostra como o algoritmo Compas classificava acusados brancos e negros e decidia se eles seriam ou não presos por uma infração futura. Esses dados coletados pela ProPublica,

do Condado de Broward, na Flórida, são mostrados na Tabela 6.1. As colunas representam as pessoas classificadas em risco mais elevado e risco menos elevado pelo algoritmo Compas. As linhas representam o número de pessoas que tiveram e que não tiveram reincidência.

Você devia olhar com atenção essa tabela e decidir se acredita ou não que o algoritmo é parcial. Primeiro, compare quantos acusados brancos e negros estão classificados como risco mais elevado. Há um total de 2.174 negros classificados como alto risco de um total de 3.615 acusados negros. Logo, a probabilidade de um acusado negro ser classificado como risco mais elevado é $2.174/3.615 = 60\%$. Se você fizer um cálculo análogo para os acusados brancos, encontrará que a probabilidade de ser classificado como risco mais elevado é somente 34,8%. Desse modo, negros têm uma maior probabilidade de serem classificados como risco mais elevado do que brancos.

Tabela 6.1 Análise detalhada das avaliações de risco baseadas no algoritmo Compas (colunas) e se a pessoa voltou ou não a cometer um crime nos dois anos seguintes à avaliação (linhas). Os detalhes de como os riscos “mais elevado” e “menos elevado” foram definidos e outros detalhes podem ser encontrados na análise da ProPublica.⁵

<i>Acusados negros</i>	<i>Risco mais elevado</i>	<i>Risco menos elevado</i>	<i>Total</i>
Reincidiu	1.369	532	1.901
Não reincidiu	805	990	1.714
Total	2.174	1.522	3.615

<i>Acusados brancos</i>	<i>Risco mais elevado</i>	<i>Risco menos elevado</i>	<i>Total</i>
Reincidiu	505	461	966
Não reincidiu	349	1.139	1.488
Total	854	1.600	2.454

Por si só, essa diferença não significa que o algoritmo seja parcial, porque a proporção de reincidentes varia entre os acusados negros e brancos: 52,9% dos acusados negros foram presos por causa de outra infração no período de dois anos, e 37,9% dos acusados brancos foram presos por causa de outra infração. Essa diferença na proporção total daqueles classificados como risco mais elevado ou menos elevado não foi a base da crítica ao algoritmo feita pela ProPublica. Julia e seus colegas admitiram que havia uma taxa maior de reincidência para acusados negros do que para acusados brancos, então questionaram que tipos de erros o algoritmo cometeu.

Quando algoritmos são avaliados, geralmente é útil pensar em termos de “falsos positivos” e “falsos negativos”. Para o algoritmo Compas, um falso positivo é quando uma pessoa que não vai cometer um crime no futuro é classificada como alto risco, i.e., a previsão do modelo foi positiva, mas errada (falsa). A fração de falso positivo é o número de acusados de risco mais elevado não reincidentes dividido pelo número total de pessoas que não reincidiram. Para acusados negros, ela corresponde a 805 dividido por 1.714, que dá 46,9%. Para acusados brancos, corresponde a 23,5%. Os acusados negros tiveram uma proporção bem maior de falsos positivos do que os acusados brancos.

Se você for preso e um juiz usar um algoritmo para ajudar a avaliá-lo, a pior coisa que pode acontecer é você receber um resultado de falso positivo. Um positivo verdadeiro é justo: o algoritmo previu que você era um risco e você era. Mas um falso positivo pode significar que lhe seja negada a liberdade condicional ou dada uma sentença maior do que você merece. Isso estava acontecendo com mais frequência com acusados negros do que com brancos. Quase metade dos acusados negros que não reincidiram tinham sido taxados como de alto risco.

Por outro lado, acusados brancos recebem mais falsos negativos, em que o algoritmo diz que uma pessoa é de baixo risco, mas ela comete crimes no futuro. Para acusados brancos, a fração de falso negativo é $461/966 = 47,7\%$ e, para acusados negros, ela é somente $532/1.901 = 28,0\%$. Para a sociedade, uma fração alta de falso negativo é um problema; isso significa que pessoas

que deveriam ter sido detidas foram devolvidas à sociedade e cometeram crimes. Quase metade das pessoas brancas que reincidiram tinham sido taxadas como de baixo risco.

Analisado em termos dessas razões de falsos positivos e falsos negativos, o algoritmo parece ser, de fato, muito ruim. Potencialmente, o Compas poderia ser usado para mandar negros para a prisão por períodos desnecessariamente longos, enquanto liberta brancos que voltam a cometer crimes.

Em resposta a essa acusação, a Northpointe argumentou que seu algoritmo devia ser julgado em relação às suas previsões serem ou não igualmente boas para pessoas brancas e negras. E elas são. Se analisarmos a primeira coluna da Tabela 6.1, vemos que 1.369 dos acusados negros classificados como alto risco, de um total de 2.174, reincidiram. Isso equivale a 63%. Para acusados brancos, são 505 de 854, ou 59,1%, de pessoas classificadas como alto risco que reincidiram. Essas razões são similares, de modo que o algoritmo está propriamente calibrado tanto para acusados brancos quanto para negros. Se um juiz recebe a pontuação de risco de uma pessoa em particular, independentemente da raça, essa pontuação reflete a probabilidade de reincidência.

Essas duas abordagens diferentes de medir a parcialidade produzem resultados contraditórios. O argumento de Julia e de seus colegas da ProPublica sobre os falsos positivos e falsos negativos é poderoso, mas o contra-argumento de Tim e de seus colegas da Northpointe sobre o algoritmo estar calibrado é sólido. A partir da mesma tabela de dados, dois grupos separados de estatísticos profissionais tiraram conclusões opostas. Nenhum dos dois cometeu erros em seus cálculos. Qual deles estava correto?

Foram necessários dois estudantes de doutorado de Stanford, Sam Corbett-Davies e Emma Pierson, trabalhando em equipe com dois pesquisadores, Avi Feller e Sharad Goel, para resolver o problema.⁶ Eles confirmaram a alegação da Northpointe de que a Tabela 6.1 mostrou que o algoritmo Compas fez previsões igualmente boas, independentemente da

raça. Depois, como os matemáticos gostam de fazer, eles chamaram atenção para um problema mais geral. Mostraram que, se um algoritmo é igualmente confiável para dois grupos, e um grupo tem maior probabilidade de reincidir do que o outro, então é impossível as taxas de falsos positivos serem as mesmas para ambos os grupos. Se acusados negros reincidem com maior frequência, então eles têm uma probabilidade maior de estarem classificados de maneira incorreta na categoria de risco mais elevado. Qualquer outro resultado significaria que o algoritmo não foi igualmente calibrado para ambas as raças, no sentido de que ele teria que fazer avaliações diferentes para acusados brancos e negros.

Para entender melhor esse assunto, façamos um experimento mental. Imagine que eu queira criar uma oferta de emprego on-line no Facebook para contratar um programador para meu grupo de pesquisa. Essa é uma tarefa simples. Basta criar uma oferta de emprego como uma publicação na página do Facebook do meu grupo de pesquisa. Depois, eu cliço no botão “impulsionar publicação”, o que facilita o direcionamento do meu anúncio.⁷ Ao usar o recurso “criar público”, consigo encontrar pessoas interessadas em cachorros, veteranos de guerra, jogadores de videogame e proprietários de motocicletas. Consigo encontrar passatempos como teatro, dança e tocar violão.

O Facebook não tem uma opção que me permita selecionar em separado homens e mulheres, não que eu ache que ele deveria ter, mas estou ciente de que, devido a escolhas educacionais diferentes na escola e na universidade, observamos um número maior de homens interessados em programação do que de mulheres. Podemos supor, a título de argumentação, que de uma população de 1.000 mulheres, 125 estão interessadas em um emprego de programador, ao passo que de uma população de 1.000 homens, 250 estão interessados.

Ao planejar meu anúncio, decidi clicar em algumas opções que acho que vão atrair programadores de computador: “jogadores de RPG”, “filmes de ficção científica” e “mangá”. Isso deve bastar. Lembrando da época em que eu era um estudante de ciência da computação, sei que muitos programadores

adoram essas atividades. Desse modo, vou atrair bons candidatos e não terei que gastar dinheiro fazendo propaganda para pessoas que não estão interessadas no emprego.

Eu finalizo meu anúncio e espero.

Depois de um dia, o Facebook mostrou meu anúncio para 500 pessoas: 100 mulheres e 400 homens.

Quando eu lhe mostro o resultado, você fica indignado. “Sua campanha publicitária é parcial”, você me diz. “RPG? Sci-fi? Aquelas opções nas quais você clicou não apenas atraem nerds da computação, mas também atraem tipicamente mais homens do que mulheres. Seu algoritmo é injusto!”

“Mas veja bem”, eu digo. “Fiz as estatísticas. Meu algoritmo é imparcial.” Pego a Tabela 6.2 e lhe mostro. Explico tudo passo a passo, no meu tom mais condescendente, superior: “Das 100 mulheres para as quais o algoritmo mostrou meu anúncio, 50 estavam interessadas no emprego e teriam se candidatado. Dos 400 homens que viram o anúncio, 200 ficaram interessados. Assim, o anúncio estava igualmente relevante para homens que o viram assim como para mulheres que o viram.”⁸

“Mas o número de homens que o viram é quatro vezes maior!”, você grita, frustrado com minha numerologia perversa, “e você sabia desde o início que pelo menos metade dos homens e das mulheres poderiam se interessar pelo emprego. Você está intensificando ainda mais a parcialidade inerente à sociedade.”

Tabela 6.2 Análise detalhada dos homens e das mulheres aos quais foi mostrada minha campanha (experimento idealizado) do Facebook.

<i>Mulheres</i>	<i>Anúncio mostrado</i>	<i>Não mostrado</i>	<i>Total</i>
Interessadas	50	75	125
Não interessadas	50	825	875
	100	900	1.000

<i>Homens</i>	<i>Anúncio mostrado</i>	<i>Não mostrado</i>	<i>Total</i>
Interessados	200	50	250
Não interessados	200	550	750
	400	600	1.000

Você está certo, é claro. Não é justo eu ter criado um anúncio que foi mostrado para um número quatro vezes maior de homens do que de mulheres. Eu tinha usado a mesma lógica que a Northpointe usou para justificar seu algoritmo. Usei a definição calibrada de imparcialidade: a proporção de indivíduos, cuja previsão de interesse no emprego foi correta, é a mesma para ambos os grupos. Foi isso que Tim Brennan argumentou quando mostrou que o resultado criminal para acusados é o mesmo para ambos os grupos. Minha campanha publicitária estimulou a eliminação de parcialidade na calibração.

Continue comigo. Você seleciona outras opções no algoritmo de anúncio do Facebook. Dessa vez, o rodamos juntos e examinamos os resultados (mostrados na Tabela 6.3). Agora o Facebook mostra o anúncio para 100 mulheres com um interesse em potencial em se candidatar para o emprego e para 200 homens que podem estar interessados. A razão 100 para 200 reflete a proporção intrínseca do total de homens e mulheres interessados no emprego (125 para 250). A proporção de falsos negativos (um para cinco) é a mesma para homens e mulheres.

Aqui vai um detalhe ao qual, mesmo que eu tenha aceitado sua maneira de fazer as coisas, não posso deixar de chamar a atenção. Somente um terço das mulheres que viram o anúncio se interessaram pelo emprego, enquanto metade dos homens que viram o anúncio ficaram interessados. Além do mais, se considerarmos os indivíduos para os quais o anúncio não foi mostrado, poderá ser dito que discriminamos os homens. Um de 11 homens que não viram o anúncio estava interessado no emprego, enquanto somente uma de 27 mulheres que não viram o anúncio estava interessada. Nosso

novo algoritmo tem uma parcialidade na calibração que favorece as mulheres.

Tabela 6.3 Análise detalhada revisada dos homens e das mulheres para os quais foi mostrada minha campanha (experimento mental) do Facebook.

<i>Mulheres</i>	<i>Anúncio mostrado</i>	<i>Não mostrado</i>	<i>Total</i>
Interessadas	100	25	125
Não interessadas	200	675	875
	300	700	1.000

<i>Homens</i>	<i>Anúncio mostrado</i>	<i>Não mostrado</i>	<i>Total</i>
Interessados	200	50	250
Não interessados	200	550	750
	400	600	1.000

A injustiça é como aquele jogo do parque de diversões em que a toupeira fica saindo de seu buraco em lugares diferentes. Você martela uma em um buraco e outra aparece em outro lugar. Tente você mesmo. Faça duas tabelas vazias 2x2 e tente distribuir 1.000 mulheres (das quais 125 estão interessadas no emprego) e 1.000 homens (dos quais 250 estão interessados no emprego) nas quatro posições de um modo totalmente imparcial. Você não vai conseguir. É impossível manter os dois grupos calibrados com razões iguais de falsos positivos e falsos negativos para homens e mulheres. Haverá sempre um grupo discriminado.

A beleza da matemática é que podemos demonstrar resultados gerais. Foi exatamente isso que os cientistas da computação da Cornell Jon Kleinberg e Manish Raghavan, juntamente com o economista de Harvard Sendhil Mullainathan, fizeram com pares de tabelas de frequência 2x2,

parecidas com as Tabelas 6.2 e 6.3. Escolhi combinações específicas de números em meu exemplo, mas Jon, Manish e Sendhil mostraram que, em geral, é impossível eliminar a parcialidade na calibração e, ao mesmo tempo, ter taxas de falsos positivos e falsos negativos idênticas para dois grupos.⁹ Esse resultado se aplica independentemente dos números que inserimos nas tabelas, com apenas uma exceção notável: no caso de as características intrínsecas dos grupos serem exatamente iguais. Assim, somente se as taxas de reincidência fossem idênticas para acusados negros e brancos do Condado de Broward, na Flórida, ou se a mesma quantidade de mulheres e de homens estudassem programação de computadores, poderíamos ter a esperança de criar algoritmos completamente imparciais. Como o mundo em que vivemos não é igualitário de todas as maneiras possíveis, então não podemos esperar que nossos algoritmos sejam completamente justos.

Não existe uma equação para a justiça. A justiça é algo humano. É algo que sentimos. O que sinto é que você foi correto quando mudou meu algoritmo de propaganda. Instintivamente, eu prefiro a Tabela 6.3 à Tabela 6.2 para uma campanha publicitária. Ao tentar encontrar a melhor pessoa para uma oferta de emprego, parece errado criar um anúncio que produza, desproporcionalmente, mais candidatos do sexo masculino do que do feminino. Na minha opinião, parece certo que deveríamos investir tempo na construção de algoritmos que sejam mais eficazes ao encontrar mulheres qualificadas para um emprego de programação, mesmo que isso signifique que nosso algoritmo não seja igualmente eficaz ao encontrar homens.

Também achei que Tim Brennan e os outros criadores do Compas estavam errados ao enfatizar a eliminação da parcialidade na calibração de suas previsões. Se a Northpointe pudesse criar um algoritmo que fosse mais eficaz ao identificar se um negro era de risco mais elevado ou não, mesmo que o algoritmo não funcionasse de modo tão eficaz para brancos, eu não diria que se trata de discriminação. O algoritmo estaria abordando um problema importante da sociedade.

Durante minha investigação no banco de dados da ProPublica, achei uma pista interessante que poderia ajudar a criar um algoritmo com uma

taxa de falso positivo melhor. Nos debates sobre o algoritmo Compas, muito pouco foi abordado sobre a razão principal de os acusados negros do Condado de Broward reincidirem com mais frequência do que os acusados brancos. A razão é bem simples, na verdade. Os acusados negros são normalmente mais jovens quando são presos.¹⁰ E jovens, de um modo geral, são mais propensos a reincidir.¹¹ Assim, se a Northpointe pudesse encontrar uma maneira de identificar jovens que, apesar de terem sido presos por um crime, são menos propensos a reincidir no futuro, então a maioria de nós concordaria que esta seria uma coisa boa. Tal método criaria inadvertidamente uma parcialidade na calibração entre brancos e negros; só porque acusados negros são mais jovens do que acusados brancos, um modelo que melhorasse sua performance para jovens teria, em média, um desempenho melhor para acusados negros.

A pergunta que eu queria fazer para Tim era se é realmente tão importante calibrar seu algoritmo, quando, em vez disso, ele poderia estar pensando em maneiras de manter jovens negros e jovens negras — que provavelmente só cometeram um erro tolo — longe da cadeia.

Poucos dias após terminar minha análise, consegui marcar uma entrevista com Tim e lhe perguntei o que ele achava da minha ideia. Ele escutou com paciência e concordou que idade, juntamente com histórico criminal e uso de drogas, são os fatores mais importantes na previsão de reincidência. Mas ele enfatizou que nos Estados Unidos existem “exigências constitucionais que demandam equidade entre raças”. De acordo com as regras da Suprema Corte, os modelos devem ser igualmente precisos para todos os grupos (no sentido de não haver parcialidade na calibração), a menos que haja uma preocupação pública muito grande sobre um assunto em particular. Então ele e seus colegas estavam sempre “andando na corda bamba” ao tentar melhorar a acurácia e seguir essas exigências.

Tim está certo de que os testes estatísticos provaram que seu modelo é imparcial e citou vários relatórios independentes que confirmam essa afirmação.¹² Ele me disse que o artigo da ProPublica fez as pessoas pensarem mais criticamente, mas que também foram uma distração para um debate

mais importante: o uso de métodos estatísticos rigorosos em condenações. “Se os níveis de acurácia de juízes foram levados em conta, então as avaliações de risco estão bem à frente dos julgamentos humanos, especialmente em relação aos erros de falsos positivos que podem [desproporcionalmente] afetar acusados negros”, disse ele.

Antes do estudo da ProPublica sobre o uso de algoritmos em condenações criminais, Jen Skeem, pesquisadora da Escola Goldman de Políticas Públicas de Berkeley, tinha conduzido uma avaliação abrangente sobre um algoritmo de condenação chamado PCRA. Ela concluiu que ele era igualmente calibrado para acusados negros e brancos e que não deveria ser rotulado de parcial. “Essas questões sobre parcialidade não são novidade”, ela me disse, “se tornaram populares agora para protestar contra o ‘algoritmo tendencioso.’”

Jen me disse que a questão mais importante é normalmente subestimada: “Como a ‘parcialidade’ se compara às práticas existentes?” Essa é a questão que ela está pesquisando atualmente.

Começo a perceber quão difícil foi decidir quem é bom e quem é mau nessa história. Meus argumentos para remover a parcialidade dos algoritmos foram baseados nas minhas experiências pessoais e nos meus valores. Mesmo que meu ponto de vista esteja certo no sentido moral, ele não se enquadra numa prova matemática válida. Tudo o que a matemática me mostrou é que não existe uma equação para justiça. É claro que Jen e Tim tinham o mesmo entusiasmo em usar algoritmos para o bem que Julia Angwin, Cathy O’Neil e Amit Datta, que conhecemos no capítulo 2. Todos eles estavam tentando fazer a coisa certa.

Sempre que recorremos à matemática para decidir a coisa certa a se fazer, ela nos dá a mesma resposta: a justiça não deriva apenas da lógica. Existem vários outros exemplos de problemas relacionados à dificuldade de se definir justiça presentes na história da matemática. O “teorema da impossibilidade”, de Kenneth Arrow, nos diz que não existe um sistema para escolher entre três candidatos políticos em que todas as preferências dos eleitores estejam igualmente representadas.¹³ O livro *Equity*, de Peyton

Young, que usa a teoria de jogos para tratar desse tema, é, como o próprio autor admite, “uma pilha de exemplos que ilustram por que a equidade não pode ser reduzida a soluções simples e abrangentes”.¹⁴ Em seu trabalho “Justice through Awareness”, de 2012, Cynthia Dwork e seus colegas procuram investigar como melhor equilibrar ação afirmativa em grupos com justiça ao indivíduo.¹⁵ Foi o que aconteceu também no trabalho de Jon Kleinberg e seus colegas sobre parcialidade, em que seus autores, ao fazerem os cálculos, encontraram paradoxos em vez de certeza racional.

Lembrei-me de um lema, outrora tão orgulhosamente proclamado por googlers: “Não seja mau.” Agora ele não é mais usado com tanta frequência na empresa. Teria o Google abandonado seu axioma depois que um de seus matemáticos descobriu que não existe uma equação que os permitisse seguramente evitar a perversidade?

Podemos nos esforçar ao máximo, mas nunca teremos certeza absoluta de estarmos ou não fazendo a coisa certa.

OS ALQUIMISTAS DOS DADOS

Vários pesquisadores e ativistas com os quais conversei até esse momento tinham certeza de uma coisa: algoritmos são espertos e estão ficando cada vez mais espertos. Os algoritmos pensam em centenas de dimensões, processando grandes quantidades de dados e aprendendo sobre nosso comportamento.

Essas percepções eram igualmente frequentes tanto entre aqueles com visões mais utópicas, como o criador do algoritmo Compas, Tim Brennan, que viu um futuro no qual os algoritmos nos ajudam a tomar decisões críticas, quanto entre aqueles com visões mais distópicas, como os blogueiros furiosos com a Cambridge Analytica. Ambos os lados acreditavam que os computadores estavam nos superando atualmente ou que nos superariam brevemente em um número grande de tarefas.

A impressão de que estamos vivenciando uma grande mudança em relação ao que os algoritmos conseguem realizar foi refletida na mídia. Os relatórios sobre o algoritmo Compas, a Cambridge Analytica e o poder da propaganda direcionada do Google e do Facebook estavam cheios de referências aos perigos em potencial da IA.

O que eu tinha descoberto até agora era totalmente diferente. Quando examinei melhor a Cambridge Analytica e personalidades políticas, encontrei limitações fundamentais na acurácia de algoritmos. Essas

limitações eram consistentes com minha própria experiência de modelar o comportamento humano. Trabalho com matemática aplicada há mais de vinte anos. Já usei modelos de regressão, redes neurais, aprendizado de máquina, análise de componentes principais e muitas outras técnicas que estavam em destaque na mídia. E, durante esse período, percebi que, quando se trata de entender o mundo ao nosso redor, os modelos matemáticos não são geralmente melhores que os humanos.

Meu ponto de vista pode ser surpreendente, uma vez que meu trabalho é usar a matemática para prever o mundo. No momento em que escrevo este livro, eu dirijo uma empresa que usa modelos para entender e prever os resultados de jogos de futebol. E coordeno um grupo acadêmico de pesquisa que usa a matemática para explicar o comportamento coletivo de humanos, formigas, peixes, pássaros e mamíferos. Estou totalmente convicto de que modelos são úteis. Assim, na minha posição, talvez não seja interessante duvidar da utilidade da matemática.

É importante, porém, que eu seja honesto. Em meu trabalho sobre futebol, conheço olheiros e analistas de times de ponta. O que me impressiona quando eu lhes apresento uma estatística sobre o poder de criação de oportunidades ou de contribuição em um jogo de um determinado jogador é o entendimento intuitivo deles sobre as razões inerentes para os números. Eu vou falar algo do tipo “o jogador X fez 34% mais assistências do que o jogador Y na mesma posição”.

O olheiro dirá: “Ok, vamos examinar as contribuições defensivas. Aí está... maior para o jogador Y. O treinador está mandando que ele seja defensivo nesta situação, reduzindo, assim, suas chances de criar oportunidades.” Enquanto os computadores são muito bons em acumular um grande número de medidas estatísticas, os humanos são muito bons em discernir as razões inerentes a essas medidas.

Um dos meus colegas especialistas em números do futebol, Garry Gelade recentemente começou a desconstruir um modelo central de análise de futebol conhecido por “expectativa de gol”. O conceito estatístico por trás da expectativa de gol é sólido. Dados são coletados em cada tentativa de gol no

futebol de alto nível: a posição, dentro ou ao redor da área, em que a tentativa foi feita; se foi de cabeça ou se foi feita com o pé; se veio de um contra-ataque ou se foi uma construção mais lenta; o nível de marcação defensiva no momento em que a tentativa foi feita e assim por diante. Depois, esses dados são usados para estabelecer o valor da expectativa de gol para cada tentativa. Tentativas feitas centralmente, dentro da grande área e com o atacante na cara do gol, recebem valores de expectativa de gol mais altos. Tentativas feitas em ângulos oblíquos ou de fora da área recebem valores de expectativa de gol mais baixos. A cada tentativa que um time faz é associado um valor de 0 (nenhuma chance de marcar) a 1 (gol garantido).¹

As expectativas de gols são úteis porque elas nos permitem avaliar como foi o desempenho de um time em jogos de poucos gols. Um jogo pode terminar 0x0, mas o time que criou muitas chances terá um total de expectativa de gols maior. Esses números têm poder preditivo: times com expectativas de gols maiores em jogos anteriores tendem a marcar mais gols reais em seus jogos subsequentes.

Quando Garry realizou sua análise, durante o verão de 2017, a expectativa de gols estava ganhando espaço na mídia tradicional. A Sky Sports e a BBC mostravam as estatísticas de expectativa de gols para as contratações de verão da Premier League inglesa; o *Guardian*, o *Telegraph* e o *Times* traziam artigos explicando o conceito e, nos Estados Unidos, a estatística era amplamente divulgada nos sites da Liga Principal de Futebol (MLS) e da Liga Nacional de Futebol Feminino (NWSL). As expectativas de gols eram gradualmente aceitas como uma maneira “objetiva” de se medir o quão bem um time havia ido.

Garry comparou a expectativa de gols a outra forma, mais humana, de se avaliar a qualidade de uma oportunidade de gol em futebol. A empresa de performance esportiva Opta coleta uma medida a que se refere como “grande oportunidade”. Estas medidas são feitas por um operador humano treinado, que assiste ao jogo e observa cuidadosamente cada tentativa de gol. Se o operador achar que a tentativa tinha uma boa chance de ser um gol, então ele a classifica como uma “grande oportunidade”. Se ele achar que não

foi uma oportunidade assim tão boa, ele a classifica como “não é uma grande oportunidade”. A comparação entre “grandes oportunidades” e “expectativa de gols” permitiu que Garry comparasse a capacidade de humanos e computadores em avaliar a qualidade das oportunidades de gol.²

Podemos avaliar a acurácia de uma “grande oportunidade” de duas formas. Em primeiro lugar, podemos observar a proporção de tentativas que não resultaram em gol, mas que foram classificadas como “grande oportunidade” pelo operador. Este valor é a fração de falso positivo, um conceito que vimos no capítulo 6. Falsos positivos são a proporção de tentativas erradas que um operador classificou como “grande oportunidade”. Em segundo lugar, podemos observar a proporção de gols que foram classificados como “grande oportunidade”. Esta é a fração de positivos verdadeiros, quando o operador fez a predição correta. Para “grandes oportunidades”, tentativas erradas foram previstas incorretamente em 7% dos casos (falsos positivos) e gols foram previstos corretamente em 53% dos casos (positivos verdadeiros).

Garry concluiu que o modelo de expectativa de gols era incapaz de reproduzir este mesmo nível de acurácia. Dependendo de como ele configurava o modelo, este gerava mais falsos positivos ou menos positivos verdadeiros, mas não conseguia igualar o poder preditivo das grandes oportunidades. Modelos de expectativa de gols usam muitos dados, mas eles (ainda) não vencem os humanos. Pode soar inicialmente impressionante que tenhamos um algoritmo para medir performance no futebol, mas o método não se sobressai comparado a um fã de futebol treinado (operadores são, tipicamente, recrutados entre entusiastas do futebol) tomando notas cada vez que um time produz uma oportunidade de marcar gol.

Garry me disse que ele realizou esta análise após ler um artigo descrevendo a expectativa de gols como o modelo “perfeito” do futebol. Se, por um lado, este exagero pode trazer benefícios potenciais a curto prazo, a longo prazo ele pode prejudicar a reputação da análise estatística no futebol. Garry trabalhou como consultor de diversos clubes, incluindo o Chelsea, o Paris Saint-Germain e o Real Madrid. Ele acredita que, pelo menos por

enquanto, em vez de substituir humanos, os modelos podem auxiliar na tomada de decisões por humanos. Ele me deu um exemplo de como a tecnologia poderia ser usada para observar goleiros durante os jogos e treinar seus posicionamentos e movimentos. A abordagem é prática e realista. Em cada aspecto do jogo, há maneiras de se utilizarem modelos, mas não existe um modelo “perfeito” do futebol.

Glenn McDonald, que trabalha no serviço de streaming de músicas Spotify, é outro expert em dados com uma atitude prática em seu trabalho. Com vários outros serviços competidores disponíveis, como a Tidal e a Apple Music, o objetivo do Spotify é ter uma vantagem sobre seus competidores em termos de sugestões para músicas novas e da criação de playlists interessantes. O Spotify alcança este objetivo ao observar nossos padrões de ouvir música. Cada recomendação que ele faz, desde a playlist “músicas de rádio” até a “feito para você”, usa um sistema de gênero musical desenvolvido por Glenn e seus colegas.

O sistema de gêneros do Spotify coloca todas as músicas como pontos em 13 dimensões, agrupando os pontos próximos como gêneros. As dimensões incluem propriedades musicais objetivas, como “volume” e “batidas por minuto”, bem como propriedades emocionais mais subjetivas, como “energia”, “valência” (tristeza) e “dançabilidade”. Nestas últimas, as medidas subjetivas são estabelecidas por meio de sessões de audição em que pessoas ouvem pares de músicas e dizem quais delas acham que é a mais triste ou mais dançável. O algoritmo aprende a diferença e classifica outras músicas apropriadamente.

Glenn criou uma visualização interativa chamada “Every Noise at Once”, que coloca todos os 1.536 gêneros musicais do Spotify numa nuvem bidimensional. De “ópera séria” na base até “re:techno” no topo; de “metal viking” no ponto mais à esquerda até “percussão africana” à direita, toda forma concebível de música é posicionada, com gêneros similares colocados próximos. É uma conquista técnica impressionante, mostrando uma maneira bem concisa de visualizar uma grande parte da herança musical mundial.

A primeira vez que falei com Glenn foi por causa de um artigo que me pediram para escrever para a revista *Economist* 1843. Antes da entrevista, eu estava um pouco nervoso sobre revelar minha opinião a respeito das sugestões do Spotify. Eu tinha usado o serviço “descobertas da semana” algumas vezes para procurar músicas novas, mas ficava frequentemente frustrado. Tenho uma tendência a gostar de músicas melancólicas, mas quando ouvi as opções sugeridas pelo Spotify, elas não tinham o mesmo efeito emocional que as minhas músicas tristes favoritas. Na verdade, as canções sugeridas costumavam ser um pouco chatas. Muitos usuários do Spotify reclamam do mesmo problema: as músicas recomendadas são versões diluídas das suas verdadeiras favoritas.

Quando eu disse a Glenn que com frequência me pego passando música atrás de música, sem me prender a nenhuma das recomendações, esperei que se mostrasse um pouco desapontado. Mas ele ficou feliz em admitir as limitações de seu algoritmo. “Não podemos esperar capturar como você se conecta pessoalmente a uma música”, ele me disse.

Glenn explicou que as playlists do Spotify funcionam melhor em festas. “É uma coisa coletiva”, ele me disse. “Somos muito bons em gerar playlists para ocasiões sociais, e o número de músicas puladas é baixo. Mas se tentamos sugerir uma música nova para você como indivíduo, então ficamos satisfeitos se você gostar de uma em dez das recomendações”. Ele está certo. Quando minha mulher e eu recebemos amigos em casa, normalmente colocamos uma playlist genérica no Spotify. Isso evita várias discussões sobre escolha musical e muitas vezes gostamos das sugestões.

Glenn me disse que o processo de fazer recomendações está longe de ser ciência pura, “metade do meu trabalho é tentar descobrir quais respostas geradas por computador fazem sentido”. Quando Glenn escolheu o título de seu trabalho, ele pediu para ser chamado de “alquimista de dados” em vez de “cientista de dados”. Ele encara seu trabalho não como uma procura por verdades absolutas sobre estilos musicais, mas como uma classificação que faça sentido para as pessoas. Esse processo requer que humanos e computadores trabalhem juntos.

Dado o amplo escopo de “Every Noise at Once”, a modéstia de Glenn ecoou fortemente em mim. Como muitos cientistas de dados com quem conversei, ele encarava seu trabalho como uma navegação em um espaço com uma dimensão grande. Mas ele foi a primeira pessoa com quem falei que admitia abertamente as dimensões irreconhecíveis e altamente pessoais da nossa mente. Ele falava sobre o que sentimos ao ouvirmos a música do nosso primeiro amor, ou a música que tocava quando dirigimos um carro pela primeira vez. Ele falava sobre músicas que nos fizeram perceber algo sobre nós mesmos e músicas que mudaram nossas atitudes com relação à homossexualidade ou ao racismo. Estas eram dimensões que Glenn admitia ser incapaz de explicar.

O conceito de alquimia de dados captura perfeitamente como a atual propaganda digital funciona. Logo após eu ter falado com Glenn, conversei com Johan Ydring, diretor de marca e performance na TUI Nordic. O grupo TUI compreende milhares de agentes de viagens e portais on-line; agencia centenas de linhas aéreas e hotéis; e lida com 20 milhões de consumidores. O trabalho de Johan é certificar-se que sua empresa faça o melhor uso de todos os dados coletados sobre seus consumidores em combinação com os dados que obtém do Facebook e outras redes sociais.

Johan descreveu suas ações como “fingindo ser esperto”. Sua equipe aparece com quatro ou cinco abordagens de propaganda direcionada a grupos específicos e as experimenta. Se uma abordagem parece funcionar, eles tentam num grupo maior.

Frequentemente as ideias mais simples são as melhores. Se um consumidor tiver feito reservas de hotel para férias na Espanha por duas vezes consecutivas, então a equipe de Johan faz com que um anúncio do Facebook apareça em seu feed de notícias pouco antes da época do ano em que este consumidor tipicamente faria sua próxima viagem de férias. O anúncio poderia sugerir Portugal, um lugar para onde o consumidor jamais tenha ido. Isso pode causar um efeito assustador no consumidor, que pode achar que o Facebook leu sua mente. Não faz muito tempo ele estava falando com um amigo sobre o Algarve e agora ele está aparecendo em sua tela. Na

verdade, ele está sentindo a ação de um truque estatístico simples. Os alquimistas de dados perceberam a época do ano em que as pessoas tipicamente agendam suas férias e descobriram uma conexão entre viagens à Espanha e a Portugal.

A maioria de nós já teve a sensação de que o Facebook ou o Google já leu nossa mente. Outro dia, antes do jantar, meu filho foi bombardeado com anúncios no YouTube sobre seu pão preferido, o menos saudável disponível no mercado. Recentemente, minha mulher comprou uma marca de chocolate pela primeira vez num mercado perto de casa e, de repente, essa mesma marca começou a aparecer como um anúncio no feed do Facebook dela.

Depois de receberem propaganda direcionada, frequentemente ouço minha família e meus amigos especularem sobre como a internet os está observando. Eles começaram a questionar se o WhatsApp pode estar vendendo suas mensagens particulares ou se seus iPhones podem estar gravando suas conversas.

Teorias da conspiração sobre empresas usando mensagens particulares dificilmente se sustentam. A explicação mais plausível é que os alquimistas de dados estão descobrindo relações estatísticas em nossos comportamentos que os ajudam a nos transformar em alvos: crianças que assistem a vídeos sobre *Minecraft* e *Overwatch* comem sanduíches à noite. É possível que minha mulher não tenha percebido que um anúncio daquela marca de chocolate já tivesse aparecido em seu Facebook.

Outra fonte importante desses anúncios “sinistros” é o redirecionamento: simplesmente esquecemos que já pesquisamos sobre uma viagem para Algarve, mas nosso navegador se lembrou e passou essa informação ao TUI, que agora lhe oferece um quarto em seu melhor hotel.

Estamos sujeitos a tamanha quantidade de propaganda e passamos longos períodos de tempo olhando para telefones e telas que, de vez em quando, parece que os anúncios leram nossa mente.

Na verdade, não é o algoritmo que é perspicaz. A inteligência vem dos alquimistas, que estão agrupando os dados de acordo com seus próprios

entendimentos sobre seus clientes. A abordagem de Johan e seus colegas é sagaz porque ela obtém resultados, algumas vezes aumentando as vendas em dez vezes, mas ela não segue uma metodologia científica bem definida. Falta rigor na abordagem deles. Johan me disse que, mesmo que tivessem um cientista de dados muito inteligente trabalhando há dez anos em modelos detalhados sobre seus clientes, ele não está convencido de que o investimento teria valido a pena. “Nossa abordagem precisa funcionar com um grande número de pessoas, e os dados não são suficientemente confiáveis para atingir grupos pequenos específicos”, ele me disse.

Ao conversar com Garry, Glenn e Johan, percebi que os algoritmos que classificam pessoas ainda precisam evoluir bastante. Como regra geral, os algoritmos simplesmente não fazem previsões tão acuradas sobre nossos comportamentos quanto outras pessoas o fariam. Os algoritmos funcionam melhor quando são usados por pessoas que entendem suas limitações.

Foi justamente quando eu estava chegando a essa conclusão que descobri outra coisa sobre o algoritmo Compas. Algo que jamais teria descoberto sozinho.

Enquanto eu estava ocupado escrevendo este livro, Julia Dressel, uma estudante de graduação de ciência da computação da Faculdade de Dartmouth, em New Hampshire, publicou uma tese notável. Ela também tinha analisado o modelo de reincidência, mas por outra perspectiva. Julia quis verificar quão bem o algoritmo funcionava em comparação com os humanos.

Para comparar humanos e algoritmos, Julia se baseou nos dados obtidos pela ProPublica sobre os criminosos do condado de Broward, na Flórida. Ela usou sexo, idade, raça, contravenções e crimes anteriores, bem como uma descrição da acusação feita, para construir parágrafos descritivos padronizados sobre os criminosos. Eles ficaram com a seguinte forma:

O(a) acusado(a) é um(a) [sexo] [raça] de [idade] anos. Ele(a) foi acusado(a) de: [descrição da acusação criminosa]. Este crime é

classificado como [grau do crime]. Ele(a) já foi condenado(a) por [número de crimes anteriores] crimes anteriores. Ele(a) tem [número de crimes juvenis] acusações de crimes juvenis e [número de contravenções juvenis] acusações de contravenções juvenis em seu histórico.

Cada uma das variáveis (raça, sexo etc.) era inserida na descrição diretamente do banco de dados do condado de Broward. Essa era toda a informação fornecida na descrição. Nenhum teste psicológico do criminoso era conduzido. Nenhuma análise estatística detalhada de crimes anteriores era executada. Nenhuma entrevista era realizada. O questionamento de Julia era se, a partir desta breve descrição, pessoas sem nenhum treinamento legal poderiam ou não prever reincidência.

Para testar sua hipótese, Julia recorreu à mesma ferramenta com que Alex Kogan tinha começado seu trabalho: o Turco Mecânico. Ela ofereceu aos trabalhadores do Turco Mecânico, todos eles com residência nos Estados Unidos, US\$1 para avaliar descrições de 50 acusados. Depois de avaliar cada descrição, eles tinham que responder, com “sim” ou “não”, à pergunta: “Você acha que essa pessoa vai cometer outro crime em dois anos?” Os trabalhadores tinham sido avisados de que receberiam um bônus de US\$5 se acertassem a resposta em pelo menos 65% dos casos. Isso deu uma motivação extra para fazerem previsões boas.

Quase metade dos trabalhadores do Turco Mecânico respondeu corretamente e ganhou o prêmio. Na média, eles acertaram 63% dos casos, de modo que pouco menos da metade deles ganhou o bônus de US\$5.

Acima de tudo, essa performance não foi significativamente pior do que a do algoritmo.³ Esses dois métodos, humano e algorítmico, foram indistinguíveis.

Esse é um resultado contundente para aqueles que defendem o uso de algoritmos em sentenças. Apesar de toda a complexidade do algoritmo — a coleta abrangente de dados; as longas entrevistas com os acusados; o uso de

análise de componentes principais e modelos de regressão; a produção de um manual operacional de 150 páginas e todo o tempo gasto para treinar juízes a usarem o algoritmo — ele produz resultados que não são melhores que aqueles obtidos por um monte de pessoas recrutadas aleatoriamente na internet. Seja lá quem forem essas pessoas, sabemos, com certeza, uma coisa sobre elas: que não têm nada melhor para fazer com seu próprio tempo do que ganhar US\$1 em um jogo de adivinhação de identidade criminal. Amadores ganharam do algoritmo.

Julia me contou que ela foi motivada a conduzir o estudo porque está assustada com as diversas maneiras com as quais a tecnologia reforça a opressão. “Os humanos rapidamente presumem que a tecnologia é objetiva e justa, então os casos em que a tecnologia não é justa são os mais perigosos”, ela me disse.

As reportagens sobre o preconceito racial do algoritmo Compas tinham, inicialmente, motivado o projeto de pesquisa de Julia. Ela queria descobrir se os humanos exibiam o mesmo preconceito que o algoritmo. A esse respeito, seus resultados reforçaram a argumentação de Tim Brennan de que seu algoritmo é justo. Os trabalhadores do Turco Mecânico deram a mesma sentença de quando lhes era apresentada a raça do acusado e de quando não lhes era apresentada a raça, e essas decisões foram compatíveis com a performance do algoritmo Compas em quase todos os aspectos. Assim, o Compas não é mais nem menos racista do que os Turcos Mecânicos.

Não, o Compas não é mais racista do que nós, mas também não é particularmente eficaz. Julia resumiu seus resultados principais de forma bem sucinta: “O que eu descobri foi que um software comercial importante, largamente usado para prever reincidência, não é mais acurado ou justo do que as previsões de pessoas com pouco ou nenhum conhecimento de justiça criminal que responderam a uma pesquisa on-line.” Concordei com ela. Meu próprio trabalho com dados mostrou que um modelo baseado somente na idade e no número de condenações anteriores tinha um nível de acurácia parecida com a do Compas.⁴ Presumivelmente, esses foram os fatores em

que os trabalhadores do Turco Mecânico se basearam para fazer seus julgamentos.

Não consigo testar de forma sistemática todos os algoritmos que são usados para medir comportamento humano e personalidade. Eu simplesmente não tenho tempo. Mas nos modelos que investiguei com mais detalhes — gols no futebol, preferências musicais, previsão de criminalidade e personalidade política —, encontrei o mesmo resultado: algoritmos, na melhor das hipóteses, equiparam-se à acurácia de humanos que realizam a mesma tarefa.

Este resultado não significa que algoritmos são inúteis. Mesmo que os algoritmos apenas se equiparem a humanos, eles ainda podem oferecer vantagens enormes em termos de velocidade. O Spotify tem milhões de usuários, e empregar um operador humano para avaliar cada uma de suas preferências musicais seria proibitivamente caro. Os alquimistas de dados da TUI usam algoritmos para garantir que escolhamos as viagens de verão que nos sejam mais adequadas.

Se um algoritmo tem o mesmo nível de performance que humanos, então os computadores ganham, pois eles têm a capacidade de lidar com dados com muito mais rapidez do que uma pessoa. Portanto, mesmo longe de serem perfeitos, modelos são certamente muito úteis.

Esse argumento de escalabilidade tem menor validade quando usado para medir a tendência de reincidência de acusados. O levantamento de dados necessário para o Compas é complexo e caro, e o número de casos processados pelo algoritmo é relativamente pequeno, de modo que o argumento de substituir humanos por máquinas é mais fraco. Existe outra grande questão sobre os algoritmos serem invasivos e o direito à privacidade dos indivíduos. Nos julgamentos do Turco Mecânico, os trabalhadores só recebiam informações públicas sobre os acusados e, mesmo assim, conseguiam dar sentenças igualmente acuradas. O processo de ser entrevistado e avaliado pode ser humilhante para muitos acusados e, ainda assim, parece que não fornece benefícios extras em termos de previsão das taxas de reincidência.

De todas as pessoas com quem conversei até agora, foi Julia quem mais me impressionou. Ela não estava trabalhando para uma empresa multinacional, como Google, Spotify ou TUI; ela não era pesquisadora de uma instituição acadêmica, como Cambridge ou Stanford, e ela não tinha o apoio de um grande meio de comunicação, como a ProPublica ou o *Guardian*. Ela era uma estudante de graduação que queria desafiar a estrutura do mundo em que vive e alcançou resultados impressionantes num curto período de tempo. São pessoas como Julia que podem impedir que sejamos dominados pelos números.

PARTE 2

INFLUENCIANDO-NOS

NATE SILVER *VERSUS* O RESTANTE DE NÓS

Recarregar, recarregar, recarregar. Quando as votações das eleições presidenciais de 2016 começaram, o site de previsões políticas FiveThirtyEight recebeu dezenas de milhões de visitas por hora. Eleitores americanos e cidadãos do mundo inteiro ficaram atualizando seus navegadores para descobrir as chances de Hilary Clinton ou Donald Trump se tornarem o próximo presidente dos Estados Unidos. O resultado esperado, dado com uma casa decimal de precisão, oscilou para cima e para baixo — 64,7%, 65,1%, 71,4% — e, dia a dia, as previsões de vitória de Hilary Clinton mudavam. Os números mais baixos chegaram a 54,6% antes do primeiro debate presidencial, os mais altos a 85,3% depois do terceiro, e agora eles finalmente tinham estacionado em 71,8%. A oscilação parou e os americanos foram às urnas.

Não tenho certeza do que os visitantes do FiveThirtyEight achavam que encontrariam quando eles atualizavam as casas decimais da previsão eleitoral. Sou um deles, e nem mesmo eu posso assegurar o que achava que queria. Algum tipo de certeza, acho.

Na manhã seguinte, toda a incerteza foi embora. Deu 100% Trump e 0% Clinton. Nenhuma casa decimal mostrada, nem exigida. A eleição tinha sido ganha pelo azarão e perdida pelo favorito.

Esta não foi a primeira falha preditiva nem a mais espetacular. Provavelmente, o pior erro de pesquisa eleitoral nos últimos anos ocorreu às vésperas da eleição geral do Reino Unido de 2015. No dia anterior à eleição, o modelo do jornal *Guardian* indicava uma disputa acirrada entre Conservadores e Trabalhistas. Na manhã seguinte, até mesmo o líder conservador David Cameron parecia surpreso com sua maioria parlamentar.

A votação seguinte da nação foi sobre a permanência do Reino Unido na União Europeia. As pesquisas mostravam resultados apertados, mas, em geral, do lado errado do resultado real do Brexit. Na ocasião da eleição geral do Reino Unido de 2017, até mesmo os pesquisadores de opinião pública pareciam confusos sobre como mostrar suas previsões. Dez dias antes da eleição, o modelo da YouGov, uma empresa de pesquisa de mercado, previu uma redução significativa na porcentagem de votos dos Conservadores comparada à eleição anterior. O modelo da YouGov contradisse outras pesquisas e, depois de ser criticada pela maneira como exibiu sua previsão na primeira página do *Times*, a YouGov admitiu estar nervosa sobre o resultado.¹ Então, quando se verificou que a YouGov tinha acertado, e a consequência foi um parlamento suspenso, ela ainda era vista como um fracasso parcial. Crescia a percepção de que os modelos não conseguiam prever eleições — que as casas decimais não faziam sentido.

Esses reveses aconteceram depois de uma década na qual os modelos estatísticos eleitorais se tornaram mais e mais comuns. Houve uma mudança na forma de divulgação de pesquisas de opinião: de jornais para sites políticos on-line, como o FiveThirtyEight, administrado por Nate Silver, e The Upshot do *New York Times*, que faz previsões probabilísticas dos resultados. Como já vimos, os algoritmos funcionam em termos de probabilidades e não em termos de resultados binários. As previsões de pesquisas não são uma exceção. Assim como nenhum algoritmo razoável declararia que é certo que uma pessoa vá cometer um crime ou ir a Portugal nas suas próximas férias, um algoritmo, mesmo um projetado para o Huffington Post, tampouco declararia que Hilary Clinton ganharia com 100% de certeza.

Um desafio para os criadores desses modelos de previsão baseados em pesquisas é que nós humanos ficamos traduzindo suas previsões probabilísticas para um sistema binário: “sim” ou “não”, “Brexit” ou “permanecer” e “Trump” ou “Clinton”. Nossas mentes preguiçosas gostam de certeza. Depois da eleição presidencial americana de 2012, quando o modelo de Nate Silver previu todos os estados americanos corretamente, ele foi declarado gênio em blogs e redes sociais. Ele era “o homem que acertou tudo”, como o *Guardian* publicou. Poucos anos depois, quando Silver deu a Trump uma chance de 5% de se tornar o candidato do Partido Republicano, o mesmo jornal o descreveu como alguém que estava cometendo “erros notórios”. Na época da eleição presidencial de 2016, para a qual o FiveThirtyEight atribuiu uma chance de 71,8% de Clinton ganhar, as redes sociais criticaram muito seus métodos. “Uma noite difícil para os analistas de números”, escreveu um colunista do *New York Times*, implicando Silver como um dos estatísticos que erraram.

Adoramos ter vilões e heróis, gênios e idiotas, de ver as coisas em preto ou branco, mas não gostamos da realidade cinzenta das probabilidades. A ascensão e queda das pesquisas não é uma exceção.

Vamos esclarecer algumas coisas desde o início. Apesar desses fracassos, os modelos usados para prever eleições são muito melhores do que decidir no cara ou coroa. Embora tenha sido atribuído ao Brexit e a Trump uma chance de vitória menor que 50% pela maioria dos modelos, essas foram exceções à regra. Na maior parte dos casos, as pesquisas — e, portanto, os modelos nos quais elas são baseadas — dão uma probabilidade que reflete a verossimilhança do resultado final. Uma chance de 28,2% de Trump ganhar não é pequena. Se eu lançar um dos meus dados de *Dungeons & Dragons* de quatro lados e obtiver um quatro, não questionaríamos meu papel de Mestre de Jogo. De fato, há evidência de que pesquisas melhores e mais frequentes estão fazendo previsões mais acuradas. A Figura 8.1 mostra a acurácia das pesquisas das eleições americanas nos últimos oitenta anos. A margem de erro em 2016 certamente não era um sinal de deterioração da acurácia.

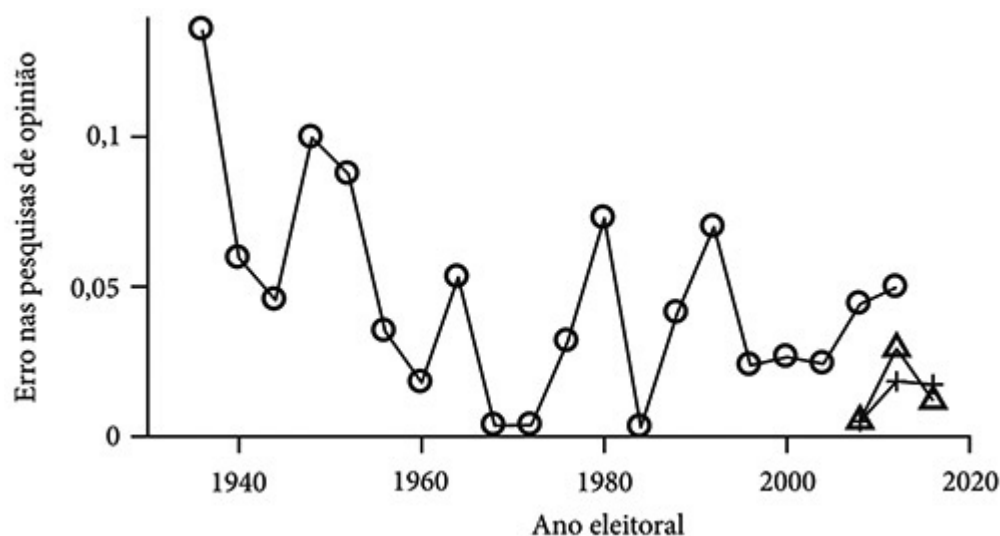


Figura 8.1 Erro entre previsões de pesquisas e resultado real sobre a margem de voto popular (relativa à soma dos votos de cada grande partido) nas eleições presidenciais dos Estados Unidos. Os círculos indicam pesquisas de opinião Gallup; triângulos indicam pesquisas da RealClearPolitics; e cruces, pesquisas do FiveThirtyEight.

Existe uma metodologia sólida e bem estabelecida por trás da maneira como as previsões eleitorais modernas são feitas. Os analistas de pesquisas de opinião tratam os resultados das eleições como curvas em forma de sino chamadas de distribuições de probabilidade, como mostra a Figura 8.2. O ponto mais alto do sino é o resultado mais provável da eleição, e a largura do sino representa nossa incerteza sobre o resultado. Um sino muito estreito indica um alto grau de certeza, e um sino largo representa um alto grau de incerteza.

A previsão envolve atualizar continuamente a forma do sino à luz de novos dados. Esse fato está ilustrado pelas curvas da Figura 8.2 de cima para baixo. Suponha que, de início, estejamos ligeiramente incertos sobre qual candidato está na frente, mas estimemos que Clinton tem +1 ponto de vantagem nas pesquisas em relação a Trump. Podemos representar essa suposição por meio de uma curva de sino razoavelmente larga, centrada em +1, como na Figura 8.2a.

Um +1 ponto de vantagem nas pesquisas significa que 50,5% das pessoas dizem que vão votar em Clinton e que 49,5% dizem que vão votar em

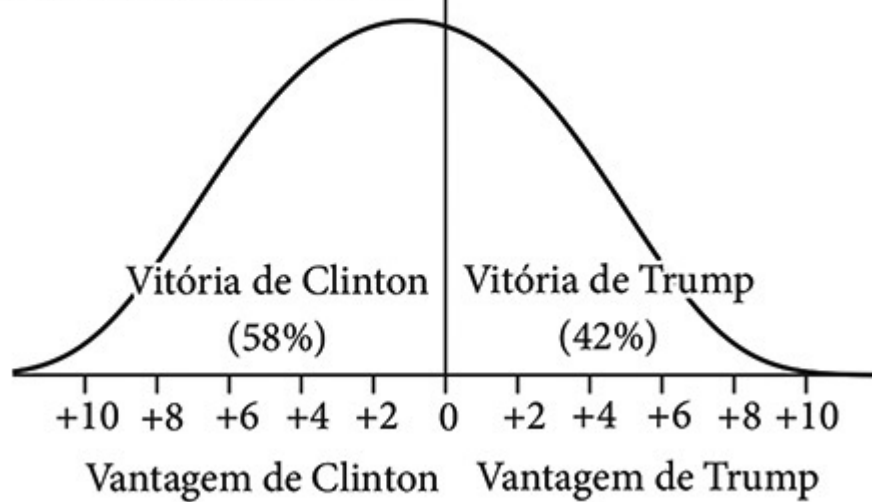
Trump. Esses números, 50,5% e 49,5%, não são as probabilidades de vitórias dos candidatos. Em vez disso, usamos a área embaixo da curva para determinar a probabilidade de que Trump ou Clinton ganharia naquele dia. Nesse caso, a área do lado de Trump equivale a 42%, e a de Clinton, a 58%. Então, apesar de achar que Clinton está na liderança, ainda damos a Trump uma chance razoável de vitória.

Suponha agora que a pesquisas revelem que Trump tem +1 ponto de vantagem em todo o país. Uma possível explicação para essa pesquisa é que Trump está genuinamente à frente e o centro do sino atual está no lugar errado. Outra explicação para essa pesquisa é que ele ainda está em desvantagem, mas a pesquisa entrevistou desproporcionalmente mais eleitores de Trump. Até mesmo as melhores pesquisas têm um grau de incerteza, mas, nesse caso, só porque elas compreenderam uma pequena parcela dos Estados Unidos e porque algumas pessoas ainda estão indecisas. Para medir essa incerteza, movemos o centro do sino para a direita, em direção à vantagem de Trump, e o alargamos (como na Figura 8.2b). Algo parecido com isso aconteceu com o modelo do FiveThirtyEight quando uma pesquisa classificada como sendo A+ (maior possível) pela Selzer & Company, em 26 de setembro de 2016, deu uma leve vantagem a Trump. Naquela época, o FiveThirtyEight tinha dado a Clinton uma chance de 58% de ganhar a eleição. Assim que ela atualizou seu modelo com os dados da pesquisa, deu a Clinton uma chance de 52%.

Nas semanas seguintes à campanha eleitoral de 2016, a maioria das pesquisas deu a Clinton uma vantagem apertada, e o sino se moveu para a esquerda e se estreitou ligeiramente. Com o tempo, as pesquisas lentamente empurraram a previsão de sua chance de vitória para mais de 70%. A essa altura, a curva em forma de sino se pareceria como a da Figura 8.2c.

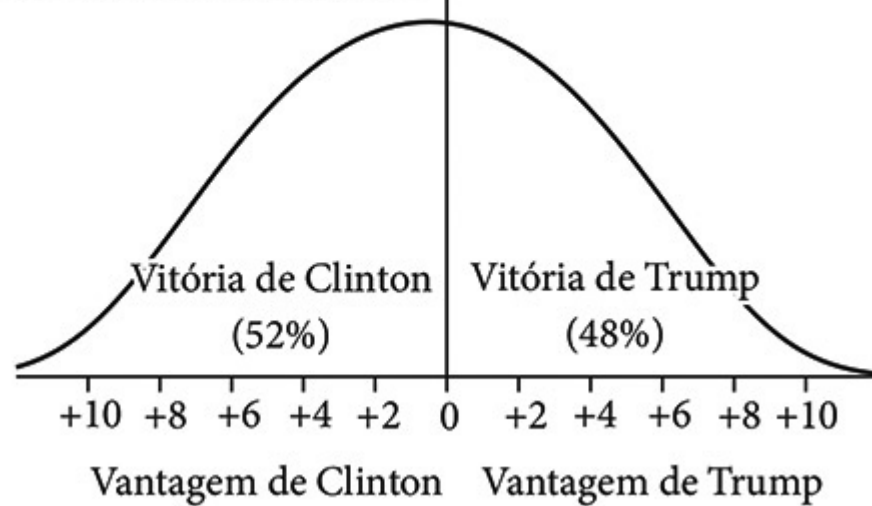
(a)

Probabilidade do resultado



(b)

Probabilidade do resultado



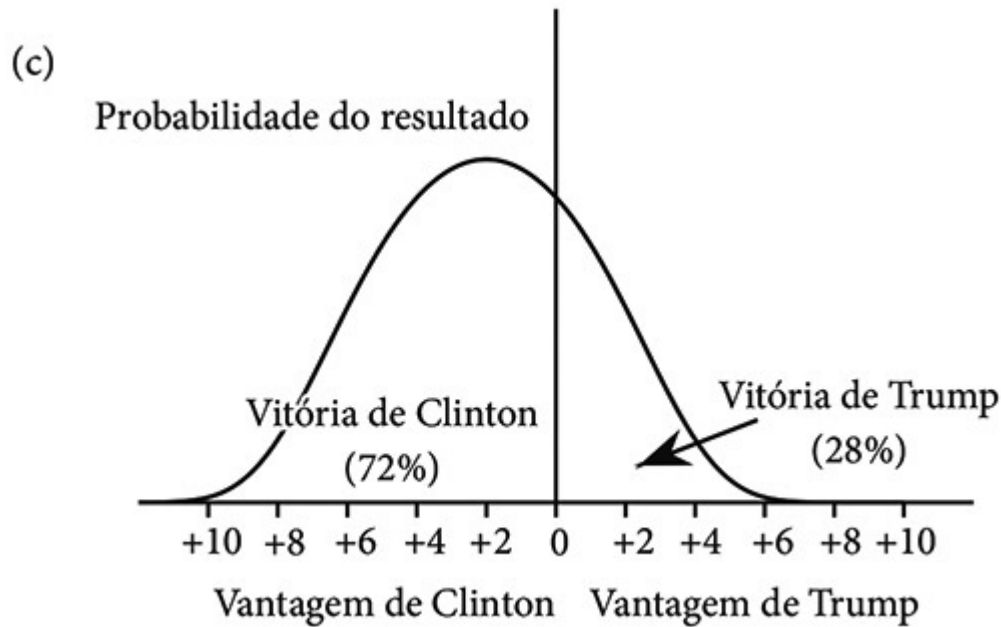


Figura 8.2a–c Três exemplos de distribuições de probabilidade de resultados em forma de sino. A altura da curva é proporcional à probabilidade de diferentes resultados (sem unidades). A área embaixo da curva, à esquerda do eixo, é a probabilidade de Clinton ganhar, enquanto a área à direita é a probabilidade de Trump ganhar. Estas curvas têm finalidade apenas ilustrativa e não representam dados reais de pesquisa.

A abordagem de distribuição de probabilidade requer a contabilidade detalhada de toda a incerteza envolvida ao se fazerem previsões. Silver e sua equipe criaram uma lista abrangente de pesquisas políticas, que foram classificadas em termos de qualidade de “A+” até “C-”, com a categoria adicional “F” para pesquisas que eles achavam que eram de dados falsos ou antiéticas. As pesquisas são avaliadas por meio dessa classificação e de quão recentemente elas foram realizadas. O FiveThirtyEight determina quanto tempo pesquisas permanecem relevantes e usa modelos de regressão para prever como as pesquisas nacionais refletem opiniões em diferentes estados. Uma vez que todos esses dados são coletados e combinados, a equipe do FiveThirtyEight simula as eleições, levando em consideração uma grande gama de erros em potencial e atualizando a curva de sino.

O número final da previsão, que todas as pessoas estavam clicando e atualizando às vésperas da eleição, é a área abaixo da curva do candidato que

aparece na liderança.

Dado o rigor por trás dessa abordagem e a quantidade de trabalho que foi gerada com as avaliações das pesquisas, Silver estava compreensivelmente preocupado com a cobertura negativa de suas previsões depois da eleição presidencial de 2016. Numa série de artigos no site do FiveThirtyEight, ele rebateu as críticas da mídia. O *New York Times* foi um foco específico de sua ira. Nas semanas que antecederam a votação, o jornal não tinha entendido que, devido às sutilezas do sistema do colégio eleitoral, Clinton tinha somente uma vantagem apertada. Um jornalista do *New York Times* escreveu que a perda de alguns pontos percentuais nas pesquisas “negaria aos Democratas uma possível vitória esmagadora e, provavelmente, daria uma vitória categórica, mas não avassaladora”.² Do ponto de vista do jornalista, a vitória de Clinton estava mais ou menos garantida. A questão era simplesmente de quanto ela seria.

O raciocínio do jornalista era uma tolice. Todas as distribuições de probabilidade da Figura 8.2 mostram uma grande variedade de cenários finais possíveis. Quando um candidato tem uma vantagem do tamanho da que Clinton tinha — correspondendo a cerca de 72% de chance de vitória — muitos resultados diferentes são possíveis, incluindo uma vitória esmagadora de Clinton e uma vitória de Trump. Não faz sentido enfatizar um resultado ou outro sem discutir as probabilidades inerentes à situação.

Quando a vitória não se materializou, o *New York Times* publicou um artigo (cuja manchete era: “Como os dados falharam ao prever uma eleição”) que afirmou que os analistas de números haviam tido uma noite difícil.³ Ele listou supostos problemas em seu próprio modelo (o modelo Upshot do NYT tinha dado uma chance de 91% de vitória para Clinton) e também na abordagem feita por Nate Silver e pelo FiveThirtyEight. O jornal estava culpando os estatísticos por não serem capazes de explicar a incerteza. Para Silver, era apenas mais um exemplo de como a mídia tem dificuldade para escrever artigos sensatos baseados em pensamentos probabilísticos.⁴

O que me espantou, considerando como o FiveThirtyEight evoluiu nos últimos dez anos, foi que o site forneceu um poderoso estudo de caso sobre as limitações dos modelos matemáticos. Nate Silver tinha sido alavancado para uma posição de autoridade. Ele tinha acumulado recursos financeiros (FiveThirtyEight pertence a ESPN) que o permitiram construir modelos sofisticados baseados em grandes quantidades de dados confiáveis. Ao ler seu livro, *O sinal e o ruído*, percebi que é um indivíduo inteligente e equilibrado, que analisou profundamente como as previsões funcionam. Ele sabe matemática e entende a relação entre dados e o mundo real. Se alguém pode criar um bom modelo de uma eleição, essa pessoa é Nate Silver.

Quando chegou o momento de analisar nosso comportamento, os algoritmos que avaliamos até agora ficam, no máximo, em pé de igualdade com os humanos. Os Turcos Mecânicos de Julia Dressel foram capazes de prever a probabilidade de reincidência a um nível similar ao de um algoritmo de ponta, usando, porém, muito menos dados; modelos de personalidade baseados em curtidas estavam longe de “nos conhecer” como indivíduos, e o Spotify estava com dificuldades para fazer sugestões de músicas tão boas quanto as dos nossos amigos.

Me perguntei se essas limitações se aplicam ao modelo do FiveThirtyEight. Dados os recursos à disposição de Nate Silver, o FiveThirtyEight era o incontestável campeão de peso pesado da previsão algorítmica. Eu queria saber se um candidato humano conseguiria derrotá-lo. Podemos nós humanos nos igualar ou mesmo superar o modelo de Nate?

Durante as primárias presidenciais, a revista *CAFE*, sediada nos Estados Unidos, contratou um especialista, Carl Diggler, com “trinta anos de experiência como jornalista político”, para fazer previsões para cada primária estadual americana. Diggler usou somente “intuição e experiência”. Ele realmente sabia como fazer isso, prevendo corretamente 20 de 22 das disputas da Super Terça-Feira.⁵ À medida que suas previsões continuaram a dar certo, ele desafiou Silver para uma competição de previsões de igual para igual. Nate não respondeu, mas Diggler persistiu mesmo assim. Ao fim das primárias, Diggler teve o mesmo nível de acurácia que o FiveThirtyEight,

com 89% de previsões corretas. Não apenas isso, ele previu o resultado do dobro de disputas que o site de Nate havia previsto. Carl Diggler foi o campeão de previsões das primárias americanas.

As previsões de Carl Diggler foram reais, mas ele não é. Trata-se de um personagem fictício. Dois jornalistas, Felix Biederman e Virgil Texas, que usaram suas próprias intuições para fazer previsões, escreveram a coluna. A ideia original deles era caçar dos entendedores de política que se pareciam com Diggler em relação a suas certezas pomposas, mas quando os jornalistas começaram a fazer previsões bem-sucedidas, eles se voltaram contra Nate. Depois da eleição, Virgil escreveu um artigo no *Washington Post* criticando a natureza enganosa do FiveThirtyEight. Ele acusou Silver de fazer previsões que não eram “refutáveis” porque elas não poderiam ser testadas, e criticaram o modo como o FiveThirtyEight pareceu restringir suas apostas usando probabilidades.

A crítica de Virgil aos métodos de Nate, baseada no sucesso de Diggler, é equivocada. Simplesmente não é verdade que o modelo de Nate não é refutável — ele pode ser testado e é o que farei nas próximas páginas. De fato, a publicação do artigo de Virgil no *Washington Post* tem uma componente importante que os psicólogos chamam de “parcialidade na seleção” e o que o guru financeiro Nassim Taleb chama de ser “enganado pela aleatoriedade”. A história só chegou ao jornal porque as previsões de Diggler deram muito certo; outros especialistas (reais ou fictícios) que falharam ao fazer previsões foram esquecidos. Ainda que seja divertido o fato de Diggler ter feito várias escolhas corretas, elas não têm validade alguma fora do mundo da sátira.

As previsões feitas pelos ditos *experts* e pelos especialistas da mídia têm sido extensivamente estudadas e comparadas com modelos estatísticos simples. Philip Tetlock, pesquisador de psicologia da Wharton School da Universidade da Pensilvânia, passou a maior parte das décadas de 1990 e 2000 estudando a acurácia de previsões feitas por *experts*. Ele resumiu o resultado mais surpreendente desse período de sua pesquisa numa única frase: “O *expert* mediano foi tão preciso quanto um chimpanzé jogando

dardos.”⁶ Pessoas como Carl Diggler, que trabalham com palpites, não superam, a longo prazo, lançamentos de cara ou coroa. Outras pessoas, possivelmente como Virgil Texas, que cuidadosamente examinam os dados antes de fazer previsões, tendem a apresentar um desempenho similar a um algoritmo estatístico simples, tal como “siga a taxa de mudança recente” ou mantenha a situação atual.

Nem Carl nem Virgil podem ser considerados adversários verdadeiros para o algoritmo do FiveThirtyEight.

A conclusão de que *experts* normalmente fracassam não foi o ponto final da pesquisa de Philip Tetlock. Ele prosseguiu com seu estudo pesquisando um grupo pequeno de pessoas que foram capazes de fazer previsões razoavelmente acuradas sobre política, economia e acontecimentos sociais. Ele chamou essas pessoas de “superprevisores” e encontrou cada vez mais pessoas com esse perfil. Eles vinham de todos os estratos sociais, mas tinham algo em comum; reuniam e ponderavam informação de uma forma que gradualmente aperfeiçoava a estimativa deles da probabilidade de ocorrência de eventos futuros. Esses superprevisores cuidadosamente ajustavam suas previsões com a chegada de novas informações. Eles eram humanos fazendo uso de raciocínio probabilístico, criando curvas em forma de sino em suas cabeças.

Os superprevisores estavam em ação nos eventos que antecederam a eleição presidencial de 2016, e, como eram uma multidão, eles se mostravam tão acurados quanto o FiveThirtyEight. Ao se fazer a média das previsões de todos os superprevisores, obtém-se uma estimativa de 24% de chance de Trump ganhar.⁷ Esse resultado foi compatível com os 28% estimados pelo FiveThirtyEight. Ao contrário de Carl Diggler, que previu 100% para Clinton, Nate Silver e os superprevisores se precaveram, e com razão.

Os superprevisores não fizeram previsões individuais para os diferentes estados americanos, o que impossibilitou uma comparação minuciosa com as previsões do FiveThirtyEight. Então, juntamente com Alex Szorkovszky, um membro do meu grupo de pesquisa, decidi investigar como um método de previsão humano se sairia nos 50 estados.

O PredictIt é um mercado on-line dirigido pela Universidade Victoria de Wellington, na Nova Zelândia. Ele permite que seus membros façam pequenas apostas sobre os resultados de eventos políticos, tais como “que partido vai ganhar em Ohio na eleição presidencial de 2016?” ou “quantos tweets de @realDonaldTrump vão mencionar CNN do meio-dia de 6/7 até meio-dia de 13/7?”. As ações de eventos são trocadas entre seus usuários diretamente, com o preço de um determinado resultado refletindo a probabilidade daquele evento. Se o preço de mercado de que Trump vai tuitar sobre a CNN mais de cinco vezes naquela semana é de 40 centavos, e se acredito que a probabilidade de ele postar seis ou mais tweets é maior que 41%, então posso fazer um investimento. Eu pago 40 centavos e, se Trump realmente postar mais de cinco CNN tweets, meu investimento gera um dólar. Se ele postar cinco ou menos, perco meu investimento.

O PredictIt permite a seus usuários negociarem com probabilidades. Não existe garantia de que todos os seus usuários sejam tão criteriosos quanto os superprevisores, mas aqueles que não considerarem as probabilidades dos eventos com cautela perderão dinheiro bem rápido. Diferentemente de casas de apostas motivadas por benefícios financeiros, o site não se esforça para tentar encorajar perdedores a apostar mais nem banir vencedores. O objetivo é fazer com que seus melhores superprevisores apostem uns contra os outros. O algoritmo PredictIt paga para pessoas que fazem boas previsões e pune aquelas que vão mal. É um modo simples e bonito de aproximar nossa sabedoria coletiva.

Em seu artigo no *Washington Post*, Virgil Texas acusou o FiveThirtyEight de fazer previsões que não podem ser refutáveis. Ele chamou previsões probabilísticas — como, por exemplo, a chance de Clinton ganhar a primária democrática é de 95% — de “afirmações instáveis”. Isso é verdade para qualquer corrida eleitoral; não tem como se refazer a votação Brexit nem as eleições presidenciais dos Estados Unidos exatamente do mesmo modo como elas foram realizadas em 2016. Mas não é verdade quando múltiplas previsões são feitas por muitos anos e em muitas eleições diferentes, como o FiveThirtyEight faz para os estados americanos. Se o

FiveThirtyEight declarar que está 95% certo em dez eventos diferentes e menos da metade desses eventos acontecerem, então devemos questionar sua metodologia. Se nove ou dez desses eventos ocorrerem, temos uma chance maior de aceitar que o método é consistente.

Uma boa maneira de visualizar a qualidade de previsões é agrupá-las com base em quão “corajosas” elas são e, então, compará-las com a quantidade de vezes em que o resultado previsto ocorre. Faço isso na Figura 8.3 para as previsões do FiveThirtyEight e as dos mercados previsores Intrade e PredictIt.⁸ As previsões corajosas são aquelas que estão nos extremos esquerdo e direito das figuras. Estas correspondem a uma previsão de que o candidato democrata vai ganhar ou perder um estado com uma certeza maior que 95%. Todas essas previsões corajosas, tanto pelo FiveThirtyEight quanto pelos mercados previsores, provaram estar corretas: o favorito ganha o estado.

Previsões mais covardes, em que a certeza da previsão fica entre 5% e 95%, são mostradas no interior das Figuras 8.3a e 8.3b. A qualidade dessas previsões pode ser medida pela distância delas até a linha pontilhada. Os círculos acima da linha indicam que as previsões tipicamente subestimam as chances democratas, enquanto os círculos abaixo da linha indicam que as chances dos democratas são superestimadas. Houve uma leve tendência em 2016, tanto pelos mercados previsores quanto pelo FiveThirtyEight, em superestimar as chances democratas em estados onde os republicanos eram os favoritos. Essa tendência não é estatisticamente significativa e pode ser razoavelmente atribuída ao acaso.

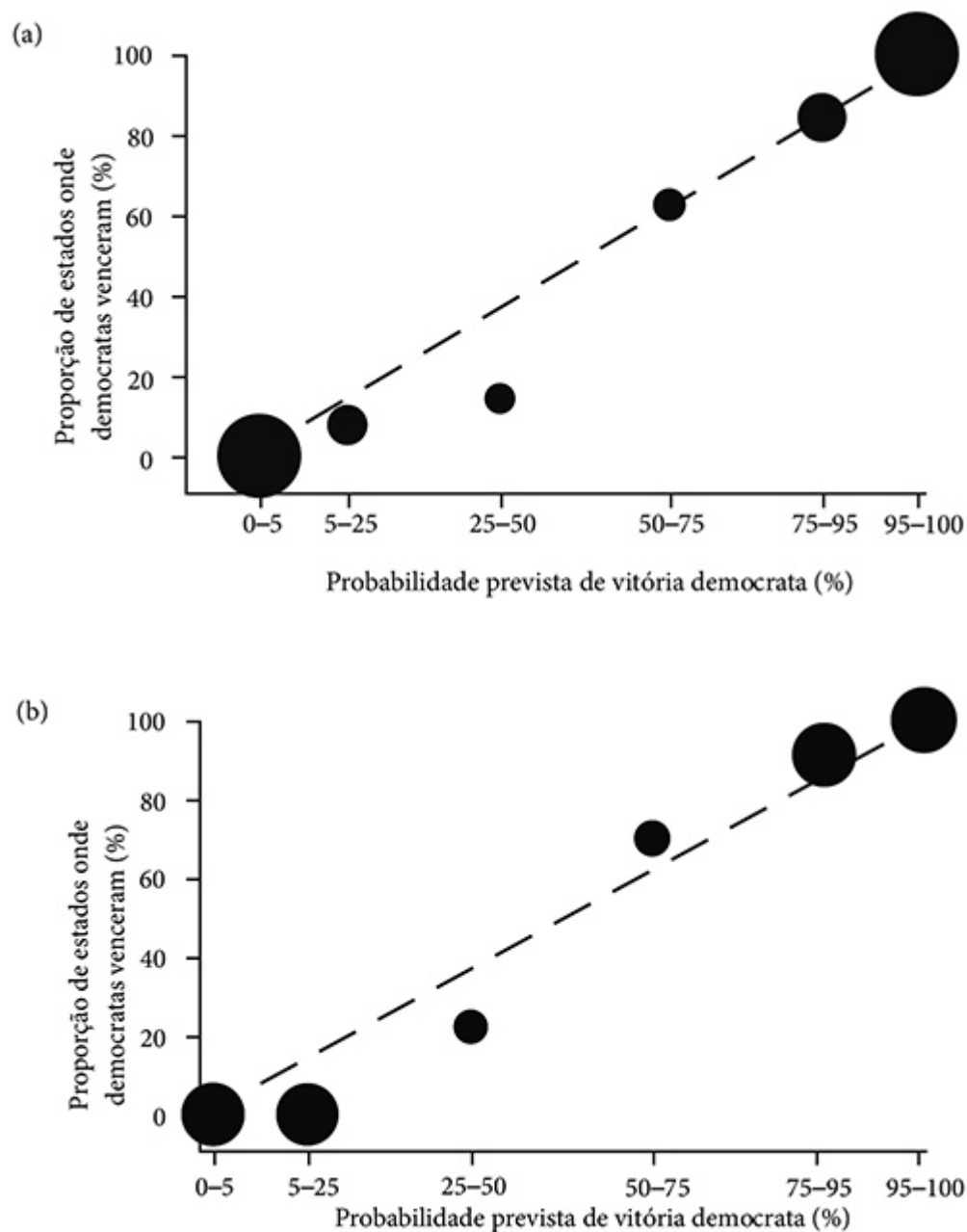


Figura 8.3 Comparação de previsões e resultados para (a) FiveThirtyEight e (b) mercados previsores Intrade/PredictIt para todos os estados dos Estados Unidos nas eleições de 2008, 2012 e 2016. Os raios dos círculos são proporcionais ao número de previsões em cada caso.

A qualidade das previsões também pode ser medida em termos dos números de previsões “covardes” e “corajosas”. Se eu tivesse declarado antes de cada eleição presidencial que aconteceu desde a Segunda Guerra Mundial que havia uma chance de 50% de um democrata ganhar, minhas previsões

teriam sido corretas, simplesmente porque cerca da metade dos presidentes foram democratas. Mas não acho que alguém diria que sou um gênio por prever cada eleição como o lançamento de uma moeda, ou seja, meio a meio.

O tamanho dos círculos da Figura 8.3 é proporcional ao número de cada previsão. Então, os círculos grandes da parte de baixo e de cima do gráfico do FiveThirtyEight mostram que eles fazem muito mais previsões “corajosas” do que “covardes”, que estão na parte do meio. Mercados de previsão são ligeiramente menos “corajosos”, especialmente no que diz respeito a fortes favoritos. Isso acontece, provavelmente, por causa do que é chamado em aposta de “favorito/tiro no escuro” — existe sempre alguém que sente prazer em fazer uma aposta comprando uma previsão de 5% de um evento muito pouco provável.⁹ Essas pessoas quase sempre (com uma probabilidade maior que 1 em 20) perdem dinheiro.

Ao calcularmos uma medida conhecida como a pontuação de Brier,¹⁰ meu colega Alex e eu descobrimos que, ao longo dos três anos eleitorais, houve muito pouca diferença no desempenho das previsões de estado a estado do modelo e da multidão humana. O FiveThirtyEight foi um pouco melhor que os mercados previsores em 2012 quando fez uma previsão bastante corajosa de que Obama era obviamente favorito. A multidão de humanos inteligentes da PredictIt e o modelo do FiveThirtyEight foram igualmente eficientes na análise da eleição presidencial de 2016. Ambos produziram opiniões similares na maioria dos estados e ambos erraram menos sobre Trump do que a impressão dada pela maior parte da mídia.

Carl Diggler e suas previsões baseadas em intuição chegaram em terceiro lugar, mas não tão atrás quanto eu acreditava que ele ficaria. A sua pontuação de Brier para as previsões estado a estado de 2016 foi de 0,084, comparado ao 0,070 do FiveThirtyEight e ao 0,075 da PredictIt (quanto menor o placar Brier, melhor). Devo dar a Virgil Texas algum crédito, ele não era assim um *expert* tão ruim quanto alguns dos chimpanzés com dardos que Philip Tetlock havia estudado.

Há um problema, entretanto, ao se comparar a PredictIt com o FiveThirtyEight. Se, por um lado, ambos fazem previsões razoáveis, por outro, não são independentes.

Se analisarmos como as duas previsões mudaram ao longo do tempo, percebemos que elas ficam bem próximas uma da outra (Figura 8.4). Isso pode ser parcialmente explicado pelo fato de os usuários da PredictIt explorarem o FiveThirtyEight. De fato, nas discussões nos fóruns dos superprevisores, a fonte mais frequente de informação era o site de Nate Silver. Contudo, o objetivo principal de um mercado previsor é que ele reúna partes de informações diferentes, ponderando-as proporcionalmente em relação à qualidade. Então, mesmo que fosse considerado como sendo de alta qualidade, é improvável que o FiveThirtyEight fosse a única fonte do mercado PredictIt. E certamente não há evidências de que as previsões da PredictIt tenham seguido, com atraso, os altos e baixos do FiveThirtyEight.

O FiveThirtyEight não usa explicitamente os dados do mercado de apostas em seu modelo. Contudo, Silver, um ex-apostador profissional, entende perfeitamente que os mercados previsores e as cotações de um apostador fazem uma análise melhor da probabilidade de um evento acontecer do que as próprias pesquisas. Ele podia perceber que os mercados não estavam muito certos sobre a vitória de Clinton. Outros modelos, do *New York Times* e do Huffington Post, que eram baseados apenas em pesquisas de opinião, estavam prevendo uma vitória de 91% e 99% para o candidato democrata, respectivamente. A equipe do FiveThirtyEight aplicou um ajuste nas pesquisas a fim de refletir a incerteza sobre o resultado, fazendo este ficar bem próximo das cotações do mercado.

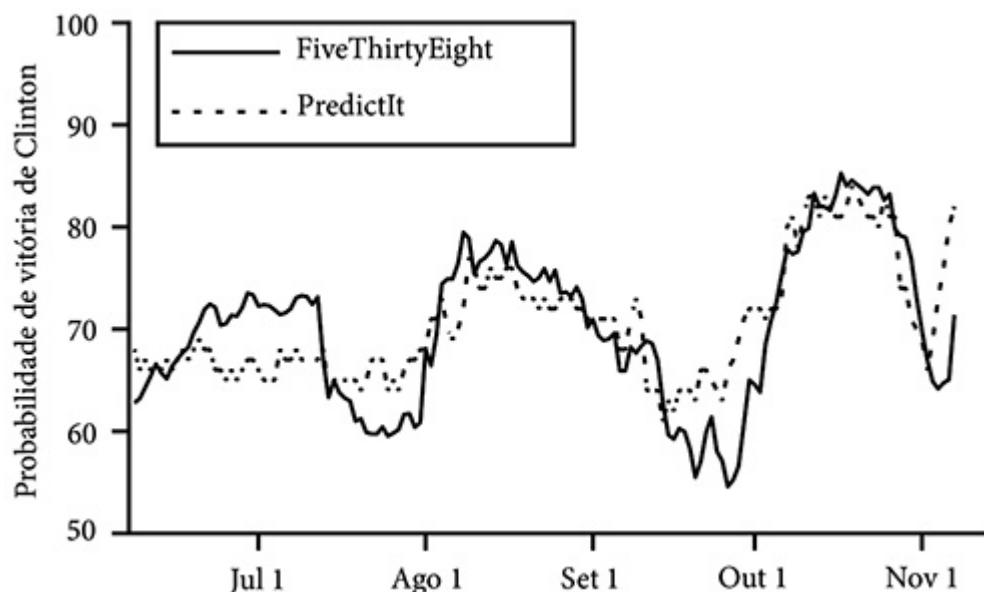


Figura 8.4 Como a probabilidade prevista da vitória de Clinton mudou durante os meses que antecederam as eleições presidenciais dos Estados Unidos, em 2016, para o FiveThirtyEight (linha sólida) e para a PredictIt (linha pontilhada).

Enquanto esse ajuste acabou sendo justificado como um modo de prever a eleição de forma menos errada que seus competidores, o ajuste fino levantou uma questão sobre o fundamento de sua abordagem. Uma ex-redatora do FiveThirtyEight, Mona Chalabi, me disse que o time de Nate usou frases como “nós simplesmente devemos ser muito cautelosos”, para expressar um entendimento comum dentro da redação de que o modelo não deveria fazer previsões tão fortes para Clinton. Eles estavam cientes de que seriam julgados depois da eleição da mesma forma, preto ou branco, que os humanos sempre julgam previsões: eles seriam ganhadores ou perdedores.

Mona, que agora é uma editora de dados do *Guardian US*, me disse: “O pior erro do FiveThirtyEight e de todas as previsões eleitorais é a crença de que haja um método que corrija todas as limitações das pesquisas. Não há.” Pesquisas acadêmicas mostraram que pesquisas de opinião são normalmente menos acuradas que previsões de mercados.¹¹ Em consequência, o FiveThirtyEight tem que arrumar um modo de melhorar suas previsões. Não existe uma metodologia estatística rigorosa para realizar

tais melhoras; eles dependem muito mais das habilidades do modelador para entender quais fatores são provavelmente mais importantes na eleição. Isso é que é a alquimia de dados: combinar as estatísticas de pesquisas de opinião com uma intuição do que está acontecendo na campanha.

Quando falei com Mona, ela enfatizou que esse assunto é muito importante: “As pesquisas são o ingrediente essencial das previsões, e as pesquisas são erradas. Então, se você tirar as pesquisas, como exatamente eles vão prever uma eleição?”

O FiveThirtyEight é uma redação quase completamente branca, sendo a maioria americana, democrata e masculina. Eles fizeram os mesmos cursos de estatística e têm a mesma concepção de mundo. Essa formação e esse treinamento significam que eles têm muito pouco insight sobre o que os eleitores pensam. Eles não conversam diretamente com as pessoas para ter uma ideia de seus sentimentos e emoções, uma abordagem que seria considerada subjetiva. Em vez disso, Mona me descreveu uma cultura em que os colegas julgam uns aos outros para determinar quão avançadas são suas técnicas matemáticas. Eles acreditavam que havia uma relação do tipo custo-benefício entre a qualidade dos resultados estatísticos e a facilidade com a qual eles podiam ser transmitidos.

Se o FiveThirtyEight tivesse oferecido um modelo puramente estatístico das pesquisas, então o histórico socioeconômico de seus estatísticos não seria relevante. Mas eles não ofereceram um modelo puramente estatístico. Tal modelo teria dado Clinton fortemente. Em vez disso, eles usaram uma combinação de suas habilidades como previsores com os números essenciais. Ambientes de trabalho consistindo em pessoas com a mesma formação e ideias são, normalmente, menos propensos a ter um bom desempenho em tarefas difíceis, como pesquisas acadêmicas e administrar uma empresa bem-sucedida.¹² É difícil para um monte de gente que tem a mesma formação identificar todos os fatores complexos envolvidos na previsão do futuro.

Não consigo ver como Silver e seu time podem derrotar o mercado eficiente da PredictIt a longo prazo. Eu mesmo não tentei prever os

resultados de eleições, mas já fiz algumas apostas no futebol (apenas para fins de pesquisas científicas, é claro). Existe uma lenda urbana do gênio matemático, talvez o Nate Silver das apostas, que descobriu uma fórmula para derrotar as casas de apostas. Diz a lenda que, se você conseguir encontrar a dica que essa pessoa pode dar, a origem da equação mágica, você ficará mais rico do que seu sonho mais ousado.

Essa lenda é pura mitologia. Não existe uma equação que preveja os resultados de eventos esportivos. A única maneira de se lucrar fazendo apostas no futebol é incorporar as cotações fornecidas pelos apostadores no seu modelo matemático. Foi o que fiz no meu último livro, *Soccermatics*, quando criei um modelo de aposta. Usei padrões estatísticos nas cotações para identificar uma pequena, porém significativa, parcialidade em como elas foram estabelecidas e explorar essas parcialidades para ganhar dinheiro. Existe matemática envolvida no processo de modelar futebol, mas um apostador que acha que consegue derrotar o mercado sem incorporar a sabedoria que já existe no âmbito do público apostador irá perder em algum momento.¹³

Esta mesma lógica do “você não consegue derrotar as casas de apostas sem fazer o jogo deles” se aplica ao trabalho de Nate. Ele admite que seus modelos esportivos não superam as cotações dos apostadores. Apostadores incorporam tanto as previsões do FiveThirtyEight como outras informações importantes ao valor de mercado. Eles sempre levam vantagem sobre um indivíduo, não importando quão inteligente este seja.

Mona aprendeu bastante com sua vivência no FiveThirtyEight, mas não do modo que ela imaginava quando começou. Ela entrou nesse emprego esperando desenvolver suas habilidades em jornalismo de dados, mas terminou entendendo que a acurácia oferecida pelo FiveThirtyEight não passava de uma ilusão. Foi ela que me chamou atenção para as casas decimais das previsões do FiveThirtyEight. Eu não tinha pensado muito sobre elas antes. Eu via probabilidades do tipo 71,8% como um número, mas tinha esquecido que elas também inferem precisão. Nas aulas de ciências da escola, aprendemos a usar algarismos significativos para representar o grau

de acurácia com o qual podemos medir uma grandeza. Por exemplo, se sabemos que 10 sacos de areia pesam entre 12,6 kg e 13,3 kg, podemos dizer que um saco de areia pesa 13 kg, com dois algarismos significativos. Todas as pesquisas de opinião têm um erro de pelo menos três pontos percentuais, e geralmente mais. Esse erro significa que, na melhor das hipóteses, as probabilidades das previsões de uma eleição deveriam ser representadas com apenas um algarismo significativo, e.g., 70%. Qualquer casa decimal a mais pode ser enganosa.

Como toda sua matemática avançada, o FiveThirtyEight está cometendo erros infantis em seus arredondamentos. Posso recomendar o site BBC Bitesize se vocês, ou eles, quiserem aprender mais sobre esse assunto.

Esse entendimento foi algo que Mona levou em consideração em seu trabalho mais recente como editora de dados na *Guardian US*. Ela usa um número pequeno de categorias ao representar dados e os ilustra de modo a enfatizar a incerteza das medidas. Ela se concentra em ordenar os dados, em vez de somente se basear nos números, e ela nunca usa casas decimais. Um de seus desenhos que mais me impressionou foi o da área de uma vaga de estacionamento e da cela solitária de uma prisão desenhados lado a lado. Não me lembro do tamanho de cada uma delas, mas a cela era muito menor que a vaga de estacionamento.

Antes de eu falar com a Mona, tenho que admitir que eu encarava a avaliação do FiveThirtyEight como um jogo. Achei que seria divertido comparar modelos de mercado para ver qual era o melhor. Mas eu estava caindo na mesma armadilha que Nate Silver. Eu estava esquecendo que o resultado da eleição presidencial dos Estados Unidos era crucial para a vida de tantas pessoas. Mona me disse que o verdadeiro perigo é que “o FiveThirtyEight possivelmente influencia o comportamento de eleitores. Os milhões de pessoas visitando o site no dia da eleição não analisam o modelo de Nate para descobrir como ele funciona. São pessoas olhando para os números de Clinton e Trump, e elas estão concluindo que Clinton vai ganhar”.

Estamos dominados por *experts* estatísticos, como Nate Silver, porque acreditamos que eles têm uma resposta melhor do que a nossa. Eles não têm. Eles podem ser melhores do que chimpanzés com dardos e eles podem derrotar por pouco um especialista (de mentira) como Carl Diggler, mas eles não derrotam nossa sabedoria coletiva. Se você estiver interessado em saber mais sobre as complexidades da criação de modelos, então recomendo muito as páginas do FiveThirtyEight (erro de arredondamento esperado). Se você só quiser verificar o número final da próxima eleição, está perdendo seu tempo. Em vez disso, use as cotações dos apostadores.

Inspirado pela ênfase de Mona na categorização por alto dos resultados diferentes, proponho um alerta obrigatório para todos os modelos de previsão, expresso de um modo que todo mundo possa entender. Poderíamos ter três categorias de desempenho:

Aleatória: essas previsões não superam os chimpanzés com dardos.

Baixa: essas previsões não superam os trabalhadores do Turco Mecânico.

Média: essas previsões não superam as chances dos apostadores.

Essas categorias cobrem os desempenhos da maior parte dos modelos matemáticos para prever eventos relacionados a humanos no esporte, política, fofoca sobre celebridade e finanças numa escala de tempo de dias, semanas e meses. Se você conhece uma exceção confiável a essa regra, por favor, me diga... Irei direto aos apostadores para confirmar o modelo.

NÓS “TAMBÉM GOSTAMOS” DA INTERNET

O algoritmo do FiveThirtyEight e o algoritmo da PredictIt eram diferentes dos que eu já havia analisado até aquele momento. Eles não somente nos classificavam. Eles interagiam conosco. O modelo do FiveThirtyEight nos influenciava. É difícil saber se, como Mona Chalabi suspeita, ele influenciou ou não o voto das pessoas, mas certamente influenciou como os americanos se sentiam em relação à eleição que se aproximava.

Interagimos com algoritmos desde o instante em que abrimos nosso computador ou ligamos nosso telefone. O Google está usando as escolhas de outras pessoas e o número de links entre as páginas para decidir quais resultados de busca nos mostrar. O Facebook usa as recomendações de nossos amigos para decidir as notícias que vemos. Reddit nos permite “votar positivamente” e “votar negativamente” em fofocas sobre celebridades. LinkedIn nos sugere pessoas que devemos conhecer no mundo profissional. Netflix e Spotify escrutinam nossas preferências cinematográficas e musicais para nos fazer sugestões. Todos esses algoritmos se baseiam na ideia de que podemos aprender seguindo as recomendações e decisões feitas por outras pessoas.

Foi realmente aqui que chegamos? Será que os algoritmos com os quais interagimos on-line realmente nos fornecem informações da melhor qualidade?

Jeff Bezos, fundador da varejista Amazon, foi a primeira pessoa a reconhecer que só queremos ver um número pequeno de escolhas relevantes ao navegarmos. Sua empresa introduziu termos como “relacionado aos itens visualizados” e “clientes que compraram este item também compraram” para nos ajudar a encontrar os produtos que queremos. A Amazon nos mostra um número pequeno de escolhas dentre milhões de opções diferentes. “Você leu *Freakonomics*? Tente *O economista clandestino* ou *Rápido e devagar: duas formas de pensar*.” “Você visualizou o último romance de Jonathan Franzen? A maior parte dos clientes acabou comprando *Uma vida pequena*, de Hanya Yanagihara.” “Frequentemente são comprados juntos: Kate Atkinson, Sebastian Faulks e William Boyd.” “Você buscou por *Soccermatics*? Tente também *The Numbers Game* e *The Mixer*.” Essas sugestões dão uma ilusão de escolha, mas os livros foram agrupados de acordo com o algoritmo da Amazon.

A razão de esse algoritmo ser tão eficiente é que ele nos entende. Quando analiso os livros sugeridos de acordo com meus autores favoritos, as recomendações são certas. Ou eu já possuo o livro ou é um que eu gostaria de ter em minhas mãos. Durante as duas horas que fiquei no site da Amazon “pesquisando” seus algoritmos, acabei colocando alguns itens em meu carrinho. O algoritmo não apenas me entendia, também entendia minha mulher e meus parentes. Fiz todas minhas compras de Natal sem precisar me levantar da cadeira. Ele até entende minha filha adolescente melhor do que eu: quando pesquisei o livro *Obsessions, Confessions and Life Lessons*, de Dodie Clark, ele sugeriu que Elise poderia gostar também de *Tartarugas até lá embaixo*, de John Green. Tenho certeza que sim.

Quando leio ficção, leio palavras de outras pessoas com minha própria voz. É uma experiência muito pessoal, uma conexão especial entre mim e o escritor. Algumas vezes, quando estou mergulhado num bom romance, acredito que nenhuma outra pessoa vai falar comigo do jeito que o escritor tem falado.

Poucas horas na Amazon dispersam completamente essa ilusão. O algoritmo usa o fato de que outras pessoas que gostam dos mesmos livros

que eu fizeram escolhas parecidas com as que eu poderia fazer. Para categorizar seus dez milhões de produtos, a Amazon cria links entre as coisas que os clientes compram juntas. Esses links são a base das sugestões que recebemos. Isso é simples, porém eficiente. Na base de clientes da Amazon, existem várias pessoas como eu, meus filhos e amigos. A verdade é que um algoritmo, desenvolvido por um pequeno grupo de pesquisadores na Califórnia, pode simplesmente escolher uma nova voz especial para eu ouvir e enviá-la, sem custos, no dia seguinte.

Eu não tenho acesso aos detalhes exatos de como o algoritmo atual da Amazon é construído: os segredos são guardados na empresa filha da Amazon, a A9, especialista em busca de produtos. O algoritmo muda com o tempo e varia para produtos diferentes, de modo que não há mais um único algoritmo “Amazon”. Existem, contudo, princípios básicos por trás do algoritmo da Amazon e como interagimos com ele, e isso pode ser capturado por um modelo matemático.

Em homenagem à Amazon, vou chamar este modelo de “também gostaram” e explicarei os passos envolvidos.¹ No meu modelo, existe um número fixo de autores. Pegarei 25 autores populares de ciências e matemática porque é divertido usar nomes conhecidos, mas os nomes não afetam o resultado do modelo. Para começar, vou supor que nenhuma compra foi feita, de modo que o primeiro cliente compra dois livros aleatoriamente e cada um dos autores tem chances iguais de ser escolhido.

A partir de então, os clientes do modelo chegam um de cada vez. Cada cliente é mais propenso a comprar livros de pares de autores que foram comprados juntos anteriormente. A princípio, este efeito é bem fraco. A regra geral é que a probabilidade de um livro de um autor ser comprado é proporcional ao número de compras anteriores mais um.² O “mais um” garante que cada livro terá uma chance de ser comprado. Imagine, por exemplo, que o primeiro cliente comprou livros de Brian Cox e Alex Bellos. A probabilidade de que um novo cliente compre Cox ou Bellos é 2 em 27, enquanto a probabilidade de que ele compre o livro de qualquer outro autor é 1 em 27.

Na Figura 9.1a, eu mostro as 20 primeiras compras numa simulação do modelo “também gostaram”. Uma linha ligando dois autores indica que um cliente comprou livros de ambos. A primeira escolha aleatória de um cliente por Cox levou a mais quatro compras conjuntas. Ian Stewart foi comprado quatro vezes. Richard Dawkins e Philip Ball venderam três livros cada. Nesta etapa, ainda não está claro qual autor é o mais popular.

Depois de 500 vendas, a situação fica bem diferente. A Figura 9.1b mostra que Steven Pinker é de longe o autor mais popular, com links fortes com Daniel Kahneman, Susan Greenfield e Philip Ball, que também estão vendendo bem. Richard Dawkins e Brian Cox ficaram para trás, e vários outros bons autores não conseguiram decolar.

Nesta simulação, alguns autores se tornaram muito populares, à medida que mais e mais conexões foram feitas entre eles, enquanto outros caíram na obscuridade. A soma das vendas dos cinco autores mais comprados é aproximadamente igual à soma das vendas de todos os outros 20 escritores juntos.

Para os autores, é aqui que reside o perigo potencial do algoritmo “também gostaram”. Os clientes do meu modelo não levam em consideração o quanto o livro é bom: eles compram livros com base nos links fornecidos pelo algoritmo. Isso significa que o número de livros vendidos por dois autores igualmente bons pode acabar em dois extremos, com um se tornando um best-seller e o outro vendendo bem menos exemplares. Apesar de os livros terem todos a mesma qualidade, alguns são best-sellers, outros são fracassos.

Kristina Lerman, uma pesquisadora do Instituto de Ciência da Informação da Universidade da Carolina do Sul, me disse que nossos cérebros adoram “também gostando”. Ela usa um princípio básico para modelar como nos comportamos on-line. Ela me disse: “Basicamente, se você supõe que pessoas são preguiçosas, então você consegue prever a maior parte do seu comportamento.”

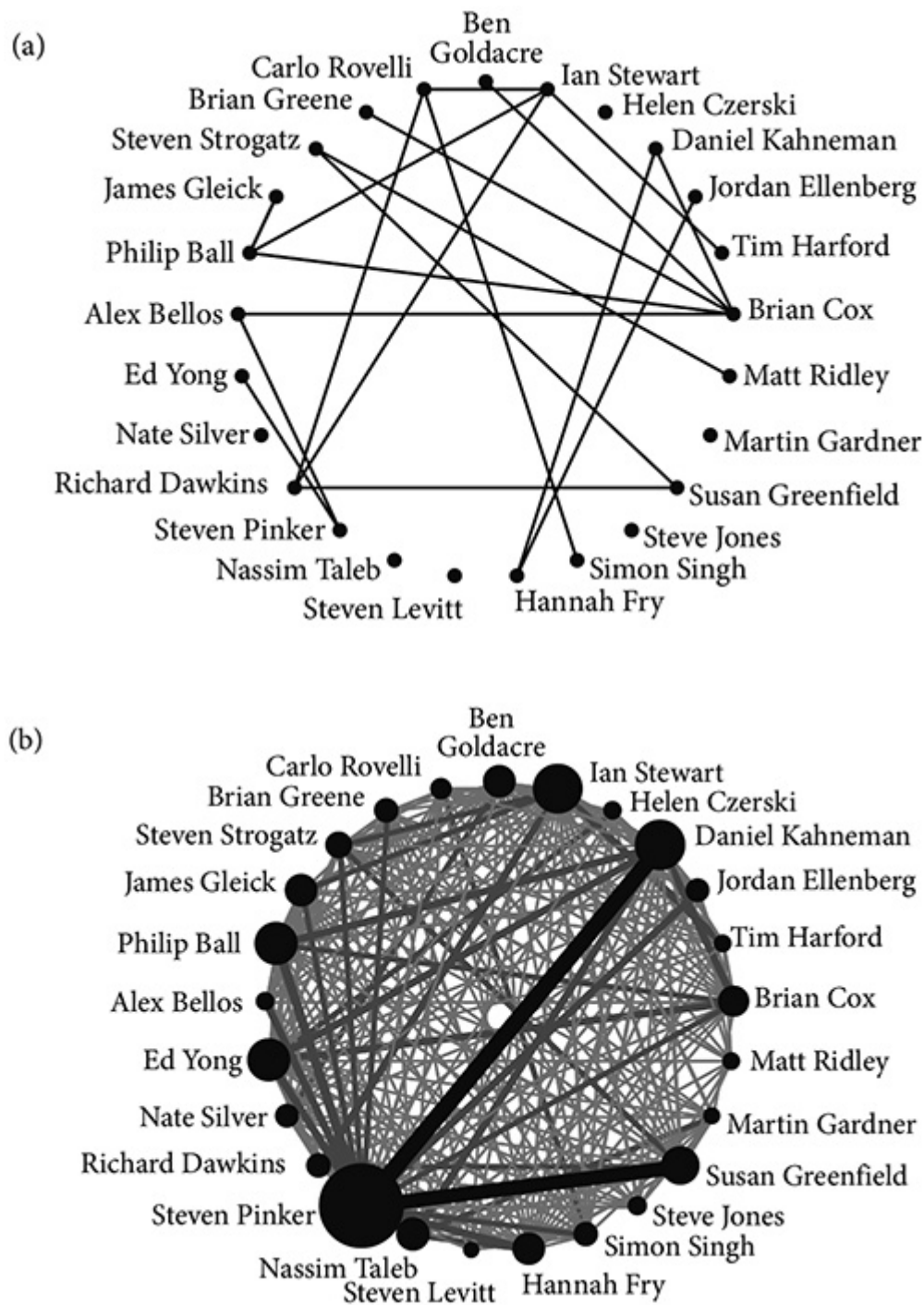


Figura 9.1 Redes mostrando as vendas de livros a partir de uma simulação do modelo “também gostaram” depois de (a) 20 vendas de livros e (b) 500 vendas de livros. As linhas mostram o número de vezes que dois autores foram comprados pelo mesmo cliente. Linhas mais espessas e mais escuras mostram compras conjuntas. O raio do círculo é proporcional ao número de vendas realizadas pelo autor.

Kristina tirou suas conclusões depois de estudar um conjunto de sites bem diversificado, de redes sociais como o Facebook e o Twitter até sites de programadores como o Stack Exchange, as vendas on-line do Yahoo, a rede acadêmica Google Acadêmico e sites de notícias on-line. Quando uma lista de artigos novos nos é oferecida, somos mais propensos a ler os que estão no topo da lista.³ Em um estudo sobre o site de perguntas e respostas sobre programação Stack Exchange, Kristina descobriu que os usuários tendem a aceitar uma resposta baseada em sua localização na página e quanto espaço ela ocupa (não o número de palavras), em vez da qualidade da resposta.⁴ Kristina me disse: “Quanto mais opções forem mostradas às pessoas neste tipo de site, menos opções elas analisam.” Quando nos apresentam muita informação, nosso cérebro decide que a melhor coisa a se fazer é simplesmente ignorar.

Kristina me disse que “também gostando” gera “mundos alternativos”, em que a popularidade on-line é decidida por várias pessoas que não estão pensando tão intensamente sobre as escolhas que estão fazendo e reforçando as decisões ruins de outras pessoas. Para entender melhor esses mundos alternativos, rodei uma nova simulação do meu modelo “também gostaram”, com os mesmos 25 autores. Uma vez que o algoritmo é probabilístico, dois resultados da simulação não são exatamente iguais. No novo resultado, mostrado na Figura 9.2, é Martin Gardner quem se destaca inicialmente e se torna um best-seller internacional. Cada rodada da simulação gera uma lista de best-seller única. Em cada mundo simulado, vendas iniciais são consolidadas e um novo escritor de sucesso de ciência popular nasce. Também gostaram cria sucessos aleatoriamente.

Não podemos reexecutar o mercado de livros real a fim de testar quanto sucesso é explicado por um golpe de sorte inicial e quanto é explicado por qualidade. Uma vez que certos autores se consagram, é impossível removê-los da história. Eu não acho que autores de ciência popular bem-sucedidos, como Ben Goldacre e Carlo Rovelli, ficariam muito empolgados se a Amazon tivesse sugerido reinicializar suas vendas de livros logo após eles terem acabado de lançar seus livros, a fim de fazer um experimento sobre

uma teoria de mundos alternativos de vendas de livros. Da mesma forma, Beyoncé, Lady Gaga e Adele não gostariam que o iTunes e o Spotify reinicializassem suas listas de músicas mais ouvidas a fim de testar uma teoria de mundos alternativos de música popular.

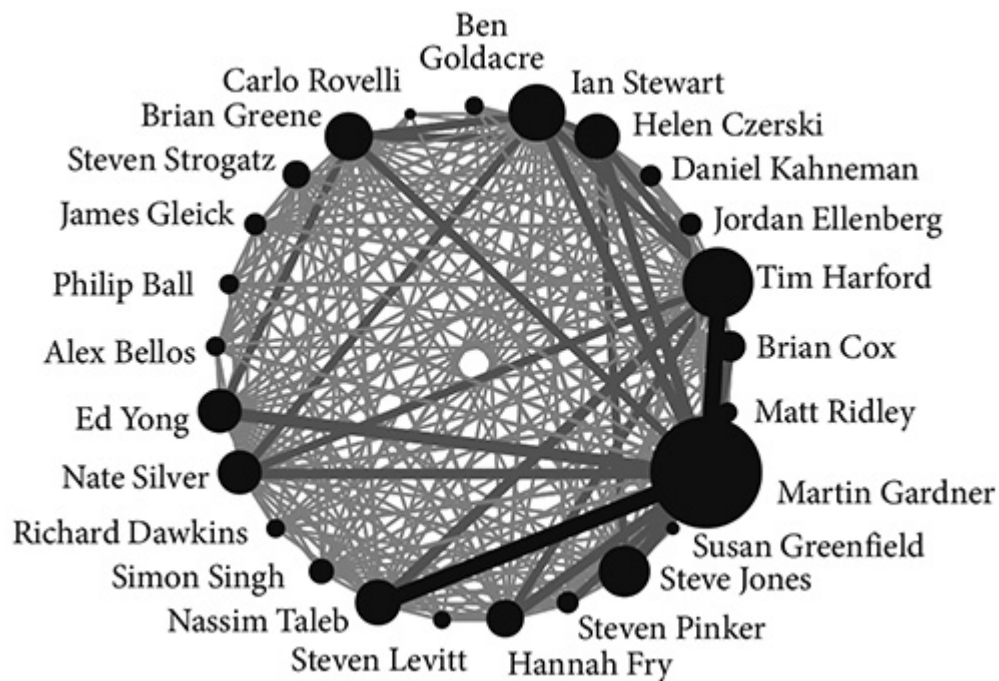


Figura 9.2 Redes mostram vendas de livros a partir de uma nova simulação do modelo “também gostaram” depois de 500 vendas de livros. As linhas mostram o número de vezes que dois autores foram comprados pelo mesmo cliente. Linhas mais espessas e escuras mostram compras conjuntas. O raio do círculo é proporcional ao número de vendas realizadas pelo autor.

Se, por um lado, não podemos reexecutar nosso próprio mundo musical, por outro, é possível criar mundos artificiais menores. O sociólogo Matthew Salganik e o matemático Duncan Watts realizaram um experimento no qual eles criaram 16 “mundos musicais” on-line diferentes em que usuários podiam escutar e fazer o download de músicas de bandas desconhecidas.⁵ As músicas eram as mesmas, mas cada mundo tinha sua própria parada musical. A parada mostrava o número de downloads de outros membros do mundo de um usuário, mas não os downloads de outros mundos. Matthew e

Duncan perceberam que as músicas do topo da parada eram cerca de dez vezes mais populares que aquelas do meio da parada. E as paradas de mundos diferentes não eram parecidas, com músicas que ocupam o topo da parada em um mundo definhando no meio da parada em outros.

Os pesquisadores também mediram o quanto as músicas eram de fato boas, por meio da observação de como ouvintes, que não tinham recebido nenhuma informação sobre a parada, as classificavam. As músicas que esses ouvintes independentes gostaram se saíram melhor nas paradas do que as músicas que os ouvintes não gostaram. Mas ainda era difícil prever qual música estaria no topo das paradas. Músicas realmente ruins nunca chegavam ao topo, mas músicas boas e razoáveis tinham chances iguais de se tornarem hits. Parece que todo mundo reconhece uma música ruim quando a escuta, mas um artista não precisa ser excepcional para fazer sucesso.

Ser “gostado” significa muito para indivíduos, negócios e meios de comunicação. Sinan Aral, da Sloan, a Escola de Negócios do MIT, começou a quantificar o impacto do “gostando”. Ele e seus colegas trabalharam em conjunto com um popular site agregador de notícias para descobrir o que acontece quando se manipulam postagens positiva ou negativamente.⁶ Acrescentar um único “voto positivo” a uma publicação logo após ela ter sido postada provocou “votos positivos” adicionais de outros usuários. Ao final da votação, o efeito do voto positivo inicial ainda podia ser sentido no número total de votos, fornecendo, na média, meio voto a mais na publicação. O efeito é pequeno, mas demonstra que a manipulação de votos funciona.

O estudo de Sinan revelou uma diferença essencial entre o algoritmo da PredictIt e o algoritmo “também gostam”. Na PredictIt existe sempre um incentivo para usuários oporem-se à sabedoria prevalecente, por causa da possibilidade de obter lucro financeiro. Usuários de agregadores de notícias não se opõem a julgamentos positivos; em vez disso, eles se tornam mais inclinados a “votar positivamente” neles mesmos. Em contrapartida, quando publicações foram manipuladas com um “voto negativo”, outros usuários

foram rápidos ao rebater com um “voto positivo”. Nesse caso, o “voto negativo” manipulado não teve um efeito global na classificação final. Nós controlamos julgamentos negativos, mas aprovamos sem questionamentos julgamentos positivos. Nosso cérebro pode ser preguiçoso, mas, pelo menos, tende a preferir positividade a negatividade.

Sinan havia concordado em não revelar explicitamente com qual site agregador de notícias ele havia trabalhado, mas, nesse momento, o site deste tipo em maior evidência é o Reddit. Erik Martin, o diretor geral do Reddit na época em que o estudo foi publicado, contou para a revista *Popular Mechanics* que o Reddit flagrou vários usuários importantes tentando manipular o site sistematicamente.⁷ O Reddit possui algorítmicos robôs (os bots) que patrulham suas páginas procurando contas falsas que publicam de maneiras não humanas. Erik disse: “Temos contramedidas e pessoas que procuram ativamente por essas coisas, uma comunidade que não tem tolerância com essa manipulação.”

O Reddit funciona bem porque humanos monitoram de perto suas discussões mais lidas, mas é impossível ter humanos monitorando e controlando cada parte da internet. E isso deixa brechas para que se explorem o “também gosto” da internet.

Localizei um velho amigo que eu sabia que poderia me dizer mais sobre essas possibilidades. Meu amigo pediu para ser identificado somente como “CCTV Simon”, sua identidade virtual, e não seu nome verdadeiro. Depois de terminar seu mestrado em Informática, Simon se tornou um pai que fica em casa, resistindo à tentação de seguir o caminho de seus colegas de mestrado, que foram trabalhar no Google e em outras empresas tecnológicas. Simon tinha algum tempo entre uma troca de fraldas e outra, e então pensou em como poderia usar esse tempo para ganhar dinheiro sem sair de casa. Foi então que ele descobriu o BlackHatWorld.

Se você estiver interessado em comprar uma câmera nova, provavelmente vai ler algumas avaliações on-line antes de realizar sua compra. Quando descobrir tudo o que precisa saber, irá à Amazon, ou a outra loja varejista líder de mercado, para realizar sua compra.

Normalmente, seu acesso à Amazon será por meio de cliques em anúncios ou seguindo links oferecidos por sites que contêm as avaliações e informações. Esses sites intermediários, conhecidos por gateways, podem solicitar um status de afiliado da Amazon. Para cada compra que se origina do gateway, a Amazon paga uma pequena comissão para esses afiliados. Para sites grandes e consagrados, essa é uma boa fonte de faturamento com propaganda. O BlackHatWorld é um fórum para afiliados que querem ganhar dinheiro, mas não desejam ter o transtorno de criar um site com conteúdo verdadeiramente útil ou interessante.

O termo *black hat* (chapéu preto) foi originalmente usado para designar hackers que entravam em sistemas de computador e os manipulavam para obter benefícios pessoais. Os afiliados do chapéu preto não invadem o Google, mas fazem tudo que podem para explorar o algoritmo de buscas do Google para ganhar dinheiro. CCTV Simon percebeu que, se ele pudesse criar um site afiliado que ficasse acima de alguns resultados de busca do Google, então um fluxo grande passaria por seu site a caminho da Amazon. Como Kristina Lerman mostrou, são nos resultados do topo que nosso cérebro preguiçoso está interessado. Com a ajuda das postagens no fórum BlackHatWorld, Simon desenvolveu uma estratégia. Ele decidiu se concentrar nas câmeras CCTV, porque elas pertencem a um ramo em expansão no Reino Unido e são suficientemente caras, de modo que a comissão poderia lhe dar dinheiro. Estudou o Google AdWords e inventou algumas frases-chaves de busca para inserir em sua página. Ao intitular uma página “Os 10 maiores erros que você pode cometer ao comprar uma câmera CCTV”, ele achou uma brecha no mercado: nenhum outro afiliado do chapéu preto havia explorado esse termo de busca em particular.

A próxima etapa é enganar o algoritmo do Google, fazendo-o acreditar que pessoas estão verdadeiramente interessadas num site afiliado. Simon chama isso de “criar o *link juice*”. No algoritmo original PageRank do Google, um site era classificado em termos do fluxo de cliques por intermédio dele, que, por sua vez, dependia da quantidade de hyperlinks de entrada e de saída que ele possuía. O algoritmo do Google se baseia no

mesmo princípio do “também gostaram” — quanto mais popular uma página é, maior a probabilidade de ela ser mostrada a outras pessoas quando elas pesquisam um assunto.⁸ À medida que uma página vai melhorando sua classificação, o fluxo do site aumenta e sua posição é melhorada ainda mais.

Afiados do chapéu preto manipulam os resultados do Google criando links múltiplos para as páginas que eles querem promover. Quando o algoritmo do Google examina uma conexão num site, acha que o site ao qual ele está relacionado é a rede central e o move para uma posição mais acima em sua lista de resultados de buscas. Uma vez que o *link juice* está fluindo e um site está indo para o topo de listas de pesquisa, ainda mais *juice* é gerado à medida que usuários reais começam a clicar. É nesse momento que os chapéus pretos ganham dinheiro, não do Google, mas dos cliques que alimentam a Amazon e de outros sites afiliados que pagam comissão.

Com o tempo, uma vez que o Google descobriu maneiras de identificar links enganosos, os chapéus pretos têm se esforçado muito para enganar o algoritmo. Atualmente, a abordagem popular é criar “redes de blogs privados”, em que uma pessoa configura dez sites diferentes sobre, por exemplo, TV widescreen. O operador contrata escritores-fantasmas para preencher o site com um texto mais ou menos sem sentido sobre o assunto. Esses dez sites se conectam então com um site afiliado, fazendo parecer que o afiliado é a autoridade máxima sobre televisores modernos. Operadores isolados do chapéu preto estão criando comunidades inteiras on-line — repletas de curtidas do Facebook e compartilhamentos do Twitter — só para enganar o algoritmo do Google.

Para dar certo, uma rede privada de blogs ou sites chapéus pretos da Amazon têm que ter algum conteúdo escrito verdadeiro. O Google usa um algoritmo de plágio para impedir que sites simplesmente copiem outros sites, e ele usa análise de linguagem automática para se certificar de que os artigos no site sigam regras básicas de gramática. Simon me disse que ele havia inicialmente comprado as câmeras CCTV e escrevera avaliações genuínas. Mas foi então que percebeu que o “Google não está nem aí se eu comprei câmeras ou não. Tudo que seu algoritmo está fazendo é checar

palavras-chaves, procurar conteúdo original, verificar se tenho algumas fotografias e medir o *juice*”. Simon sabia, de seus estudos na universidade, como o algoritmo funcionava, mas ficou impressionado em como a abordagem do Google em relação ao fluxo era grosseira. Rapidamente seu site estava atraindo centenas de milhares de visualizações.

Sites afiliados variam imensamente. Simon comparou seu próprio site com os afiliados “chapéus brancos”, que ele descreveu como “mães, donas de casa dos Estados Unidos, muito genuínas, muito pessoais, que têm fotos delas mesmas e realmente se preocupam com os produtos que apresentam”. Existe também uma área cinzenta, ocupada por sites como o HotUKDeals. Este site encoraja seus membros a compartilhar dicas sobre os maiores varejistas, dando a impressão de ser uma comunidade de caçadores de barganhas. Apesar de o site ter uma base de usuários grande e genuína, descobri que aqueles que fazem publicações no HotUKDeals também estavam sendo recrutados no BlackHatWorld a fim de fazer publicações para certos sites afiliados. Simon acreditava que a maior parte dos usuários estava alheia ao verdadeiro propósito do site: cada um dos links do HotUKDeals é um dos sites afiliados, de modo que toda dica gerava dinheiro para o dono do site.

No auge das operações de Simon, ele estava ganhando £1.000 por mês de um site contendo avaliações inventadas e dicas sem sentido. Ao navegar no site, fiquei impressionado com o estilo de escrita que ele tinha dominado: uma espécie de avaliação no estilo *Top Gear* para CCTV que realmente não dizia coisa alguma. O site faz perguntas e considerações para o leitor, tais como “quer uma câmera IP interna de baixo custo?” ou “não se esqueça de fazer seu dever de casa se você quiser ficar de olho no bebê”. Ele “explica” alguns dos jargões e fornece longas avaliações que evitam a questão de alguém no site ter de fato usado ou não a câmera.

Apesar de seu site de CCTV ainda lhe render umas duzentas libras por mês, Simon não o mantém mais nem o atualiza. Ele até considerou investir na construção de um número maior de sites afiliados, mas pensou: “Eu conseguiria colocar meus filhos na cama à noite e dizer-lhes que o papai teve

um ótimo dia de trabalho honesto?” Sua resposta foi não, e quando ele retornou à ativa, arrumou um emprego.

Procurando no Google, percebi que todas as cinco sugestões do topo para “câmera CCTV residencial” possuem links para a Amazon. Nenhuma das “avaliações” dá qualquer indicação de que quem escreveu tenha de fato utilizado os produtos. A revista *Which?* está no sétimo lugar na classificação do Google e garante possuir avaliações verdadeiras, mas elas ficam protegidas por um sistema de assinaturas. Em vigésimo lugar, o *Independent* possuía algumas boas avaliações, com links relativamente discretos para os fabricantes.

Quando interesses comerciais estão envolvidos, a confiança coletiva da Amazon e do Google pode cair de forma dramática. O retorno positivo criado pelo “também gostando” e a importância que o Google coloca no fluxo implicam que um afiliado de chapéu branco genuíno — que realmente fez avaliações sistemáticas sobre todas as TVs widescreen ou câmeras CCTV — iria rapidamente desaparecer em meio a todos os sites de chapéu preto que estão apenas inflando seu *juice*. Diferentemente do algoritmo da PredictIt, no qual o incentivo financeiro para os apostadores é fazer previsões melhores, os incentivos financeiros relacionados à compra e venda de produtos on-line residem no aumento de incerteza para os consumidores.

Com toda essa distorção on-line, foi inevitável não pensar um pouco sobre meu próprio sucesso pessoal. Como será que *Dominados pelos números* irá se sair quando for lançado? Eu não vou criar um exército de robôs para clicar nos links da Amazon ou gerar uma comunidade de pessoas recomendando o livro no site Goodreads. Será que há alguma chance de meu livro se tornar um best-seller?

Mostrei os resultados da minha simulação do “também gostaram” ao sociólogo Marc Keuschnigg. Ele havia estudado vendas de livros em detalhe, tentando encontrar o segredo de como fazer um best-seller. Ele concordou comigo que uma grande parte da explicação do sucesso ou não de um livro

está no quanto ele foi “também gostaram” na Amazon. Mas nem todos os livros e autores são iguais.

“Há uma grande diferença entre estreantes e autores estabelecidos. São exatamente os estreantes que estão mais suscetíveis ao efeito dos pares”, ele me disse. “Quando as pessoas não têm informação sobre qual livro comprar, elas pesquisam para verificar o que seus pares compraram.”

Marc estudou vendas de ficção alemã em livrarias de 2001 a 2006, pouco antes de a Amazon se tornar dominante no mercado, e descobriu que, se um novato entrava na lista dos mais vendidos, então suas vendas eram impulsionadas em 73% na semana seguinte. A visibilidade em estar nas 20 primeiras posições aumentava ainda mais as vendas.

A visibilidade em listas de mais vendidos era apenas parte da história. Outro fator que ajudou os estreantes eram as avaliações ruins na imprensa. Sim, é isso mesmo, não avaliações boas, mas ruins. Uma avaliação negativa em um jornal ou revista produzia normalmente 23% de aumento nas vendas para livros escritos por romancistas de primeira viagem. Avaliações positivas, por outro lado, não surtiam qualquer efeito. Marc me disse: “Há um alto risco de que as listas de best-sellers contenham, em sua maioria, livros medianos e de qualidade ruim.”

Para dar suporte a tal afirmação incisiva, Marc me mostrou uma análise que ele havia feito relacionando vendas a avaliações on-line. Quanto mais exemplares um livro vendia, menos estrelas ele recebia na Amazon. Esta pode muito bem ser uma forma de vingança de leitores desapontados.⁹ Eles veem um livro escalar as listas, ou aparecer em uma relação de “também gostaram”, e são persuadidos a comprá-lo. Quando percebem que o livro é chato, descontam sua frustração atribuindo-lhe uma pontuação ruim na Amazon. Parece que os leitores nunca aprendem: ignoramos as avaliações e seguimos o público. Só depois que o público se engana é que as avaliações começam a refletir a qualidade do livro.

Do meu ponto de vista, o papel do “também gostaram” na distorção da qualidade dá um gosto amargo ao sucesso. Fiquei satisfeito com meu último livro, *Soccermatics*, que vendeu razoavelmente, mas, por outro lado, recebi

uma avaliação de uma estrela na Amazon que me fez pensar no estudo de Marc. Este leitor havia comprado o livro porque tinha lido em um fórum de apostas que se tratava de um manual para ganhar dinheiro em cima das casas de apostas. Apesar de não ter sido essa a minha intenção, o avaliador havia ficado bastante desapontado. Ele (supondo o gênero) escreveu: “99% das avaliações aqui são falsas. Eu li o livro, muito, muito ruim, você irá jogar dinheiro fora com ele, você não irá aumentar suas chances de ganhar apostas lendo este livro e irá apenas encher os bolsos desse cara.”

Quando sabemos que estamos vivendo em um dentre muitos mundos alternativos, o sucesso em nosso mundo real pode parecer muito vazio.

O CONCURSO DE POPULARIDADE

Durante o verão de 2017, meu filho me fez ouvir uma música realmente ruim: um rap chamado “It’s Everyday Bro”. Acompanhado de batidas de hip-hop californianas, ela começava com uma frase que me fez passar mal: “It’s everyday bro with that Disney Channel flow” [É todo dia mano, com a cadência do Disney Channel]. Aí ela continua com exaltações sobre o número de seguidores que o artista, Jake Paul, tem no YouTube, até que ele passa a bola para sua galera “Team 10” cair dentro com frases do tipo: “My name is Nick Crompton. Yes, I can rap, but no I’m not from Compton” [Meu nome é Nick Crompton... Sim, eu sei cantar rap, mas não, não sou de Compton].

Eu sei, e meu filho me confirmou, que a música foi feita para ser um tanto irônica. Mas Jake Paul acabou revelando a extensão completa de a popularidade ter se tornado uma forma deturpada do “também gostando”. Jake ganhou fama no site de vídeos Vine e, após fazer um papel no Disney Channel, ele lançou seu canal do YouTube. Lá, podemos vê-lo dirigindo sua “Lambo” pela rua de sua escola antiga, somos levados a um tour pela sua mansão luxuosa em Beverly Hills e podemos vê-lo berrando das janelas de um hotel italiano. Em todos os seus filmes, músicas e comentários em redes sociais, ele estimula seus seguidores a “gostar” e compartilhar tudo o que ele faz.

Um aspecto original da explosão de popularidade de Jake Paul é que ele encontrou uma forma de explorar os “não gostei” do YouTube da mesma forma que os “gostei”. Um fator de vendas extraordinário de seu vídeo “It’s Everyday Bro” é que ele recebeu dois milhões de não gostei, mais do que qualquer outro vídeo jamais carregado. Algo “nunca feito antes”, como Jake Paul diria. As crianças estão assistindo ao vídeo e aos comerciais de 10 segundos que o precedem, apenas para não gostar dele. Paul se tornou popular por ser tão medíocre, tão alienado e tão desavergonhado em sua ânsia por popularidade.

Durante a segunda metade de 2017, muitos youtubers — aqueles cuja atividade on-line principal era até esse ponto filmar a si próprios jogando jogos de computador ou pregando peças em seus amigos — começaram a contratar produtores musicais e artistas de rap para ajudá-los a fazer *diss tracks* (músicas criadas para insultar uma pessoa ou grupo de pessoas) sobre outros youtubers. Uma “estrela” pioneira deste movimento é RiceGum. Ele se especializou em fazer graça dos vídeos que outras pessoas carregam, frequentemente na forma de raps, na esperança de que eles o “insultem” de volta, devolvendo tráfego ao seu canal. Muito do conteúdo dos insultos e contrainsultos se refere explicitamente a quanto são “curtidos” e quanto dinheiro eles têm. Quando RiceGum esculhambou o vídeo de Jake Paul em que ele fazia um tour em sua mansão, Jake respondeu com um discurso inflamado de como a Lamborghini de RiceGum “era alugada” e que ele “ganhava apenas US\$60.000 por dia”. Surpreendentemente, quanto mais referências esses youtubers fazem sobre serem ricos e curtidos, mais crianças os seguem, se inscrevem em seus canais, assistem a anúncios sem fim e compram seus “troços” (mercadorias).

Não há nada particularmente extraordinário ou especial sobre Jake Paul ou RiceGum. Eles são jovens que tiveram sorte de alcançar o topo de todas as listas de “tem que assistir” do YouTube. À medida que eles recebem mais “gostei” e “não gostei”, mais arrecadação com propagandas e mais vendas no iTunes, eles ganham mais atenção e fazem ainda mais sucesso. Eles são o

produto de um algoritmo que recompensa a busca por atenção e ações para chocar.

Listas de mais vendidos, prêmios e vantagens que vêm com o sucesso e contatos acumulados sempre fizeram parte de nossas vidas. Os sociólogos há muito conhecem o efeito “rico ficando mais rico”. Assim como CCTV Simon, Jake Paul e RiceGum também compreendem a importância do *click juice* e em ser “também gostaram”, eles encorajam explicitamente seu exército de fãs a se envolver. Os vídeos “próximo” do YouTube amplificam este efeito. Cada decisão de assistir, escutar ou comprar que seus fãs fazem é acompanhada por um pedacinho de informação sobre as escolhas feitas por outros. Por meio de uma série de cliques e sugestões, meninos e meninas adolescentes e pré-adolescentes criam um mundo de superestrelas cujo sucesso reflete parcialmente a qualidade de seu trabalho, mas também reflete a necessidade desesperada de se certificar de que eles não estejam passando despercebidos. Para Jake Paul, a jornada da obscuridade relativa à dominação do mundo durou apenas seis meses. Pode acontecer com qualquer um com um talento módico, mas não acontece com a maioria de nós.

Na área acadêmica, o equivalente a alcançar o topo de inscritos no YouTube é alcançar o topo nas listas de citações no serviço Google Acadêmico. Como o YouTube, o Acadêmico provê uma única e simples função: uma lista de links para artigos relacionados ao termo de busca que foi fornecido. A ordem na qual os artigos são apresentados é determinada pelo número de vezes que eles foram citados (ou referidos) por outros artigos.

Citações são essenciais na academia porque elas criam nosso discurso. Referências em um artigo e uma lista de citações em seu final mostram como o artigo contribui para um entendimento conjunto de problemas. O número de citações que um artigo atrai é uma forma muito boa de se avaliar a importância de um artigo em algum campo. Quanto mais citações um artigo tiver, melhor ele representa como os cientistas estão pensando.

Ordenar artigos por citações faz sentido, mas a abordagem do Google Acadêmico possui um efeito colateral inesperado. Quando o website começou em 2004, a revista *Nature* entrevistou o neurocientista Thomas Mrsic-Flogel. Ele lhes disse: “Eu sigo o rastro das citações e me deparo com artigos que não esperava.”¹ Em vez de visitar a biblioteca, ou mesmo as páginas de revistas científicas, ele usava os links de citação entre artigos para encontrar ideias novas. Assim como meu filho navega de um youtuber para outro, Thomas estava navegando de um cientista para outro.

Não estou julgando. Na época eu estava fazendo exatamente a mesma coisa que Thomas. Ainda faço hoje em dia. Navego pelos artigos, clicando próximo ao topo das listas, tentando entender o que está acontecendo na minha área de pesquisa e quem está publicando as melhores coisas. Assim o fazem, também, todos os meus colegas. Logo após o Acadêmico aparecer, estávamos todos viciados.

O engenheiro do Google Anurag Acharya, cocriador do Acadêmico e que continua a trabalhar no site, havia dito que seu objetivo original era “tornar os pesquisadores mundiais 10% mais eficientes”.² Este era um objetivo bastante ambicioso, mas ele já o alcançou. Ao escrever um livro como este, eu faço cerca de 20-50 buscas no Google Acadêmico por dia, uma incrível economia de tempo e esforço. Este livro e minha pesquisa seriam impossíveis sem o Acadêmico.

O que Anurag não sabia na época era que ele havia criado um algoritmo também gostei para a área acadêmica. Se um artigo é citado mais vezes, ele vai para uma posição mais alta na lista de resultados e é mais fácil de ser visto por outros cientistas. Isto significa que artigos populares são lidos e citados mais frequentemente, causando um retorno; alguns artigos sobem nas listas de citações e outros caem. Assim como livros, músicas e vídeos no YouTube, a ascensão e queda das citações de um artigo científico pode ter menos a ver com a qualidade genuína do que com pequenas diferenças em sua popularidade inicial.

Em situações em que um grande número de itens pode ser “também gostado” — como as centenas de milhares de artigos científicos que já foram

escritos — a popularidade pode ser descrita frequentemente por uma relação matemática conhecida como “lei de potência”. Para entender leis de potência, pense em um gráfico da proporção de artigos que são citados mais do que uma certa quantidade de vezes. Estamos mais acostumados a gráficos com pontos que aumentam em uma escala linear, i.e., com valores igualmente espaçados, como 1, 2, 3, 4 etc. ou 10%, 20%, 30% etc. Leis de potência são reveladas quando os dados são colocados em um gráfico log-log, em que aumentamos os espaços em potências sucessivas de um certo número. Por exemplo, as potências positivas (ou, para ser mais exato, não negativas) de 10 são 1, 10, 100, 1.000, 10.000 etc. De modo análogo, as potências negativas de 10 são sucessivamente menores: 0,1; 0,01; 0,001 etc. Nessa escala log-log, o eixo x representa quantas vezes um artigo foi citado, e o eixo y representa a proporção de artigos que foram citados este número de vezes, ou mais.

Na Figura 10.1 é mostrado um gráfico log-log para artigos científicos em 2008. Há uma relação linear (linha reta) entre a proporção de artigos e o número de citações (para artigos com mais de cerca de 10 citações). É esta linha reta que é conhecida como lei de potência.*

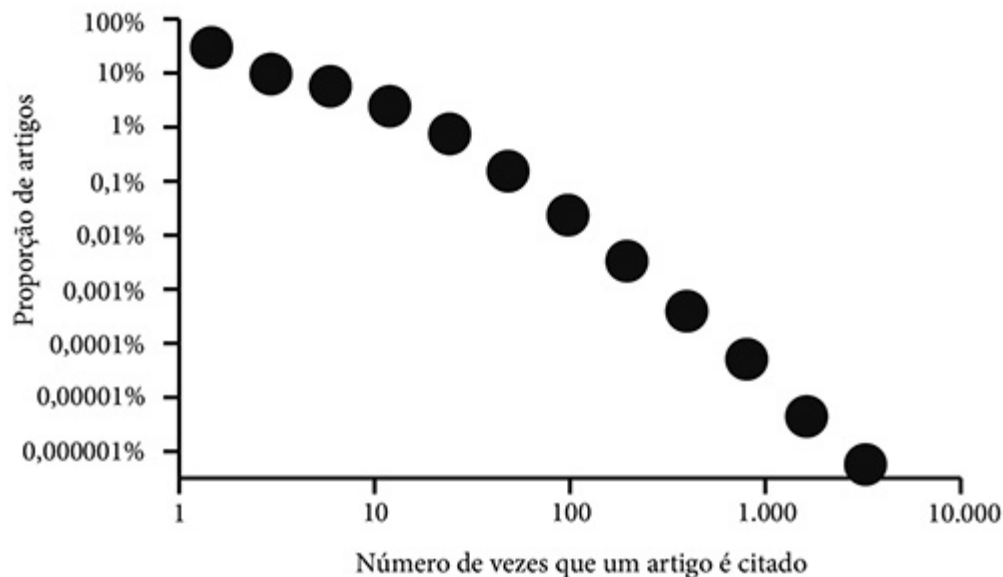


Figura 10.1 O número de vezes que um artigo é citado mostrado em função da proporção de artigos citados mais do que este mesmo número de vezes para artigos científicos em 2008. Dados coletados por Young-Ho Eom e Santo Fortunato.³

Leis de potência são um sinal de ampla desigualdade. Em 2008, 73% dos artigos científicos haviam sido citados uma vez ou menos. Um pensamento deprimente para qualquer um que tenha passado muitos meses escrevendo um artigo. Do outro lado do espectro, um em 100.000 artigos foram citados 2.000 vezes ou mais. Há muitos artigos sem sucesso, que ninguém jamais lê, e alguns poucos artigos com muito sucesso. Exatamente a mesma coisa acontece com youtubers: há vinte e tantos canais como os de Jake Paul, PewDiePie, que filma a si mesmo jogando jogos de computador, e o DudePerfect, que faz coisas impossíveis com bolas de basquete, cartas ou outros equipamentos esportivos — que têm dezenas de milhões de inscritos, enquanto centenas de milhares de canais possuem apenas um punhado.

Comparando um modelo com o gráfico log-log de dados de citações, os físicos teóricos Young-Ho Eom e Santo Fortunato mediram como a importância relativa do “também gostando” (citando artigos que já haviam sido citados muitas vezes) mudou com o tempo. A lei de potência em forma de linha reta de 2008 é mais bem explicada por um alto grau de “também gostando”, enquanto um nível menor de desigualdade observado em anos anteriores é consistente com uma maior tendência dos cientistas em tomar decisões independentes. Os anos intermediários testemunharam o surgimento do concurso de popularidade científico, em que artigos “também curtidos” decolam e artigos não curtidos são deixados para trás. A desigualdade continuou a acelerar. Em 2015, 1% dos artigos em revistas de ponta contabilizavam 17% das citações nessas revistas.

Devido ao risco de o “também gostando” dar origem a um concurso de popularidade científico, podemos imaginar que nossa comunidade tenha tomado bastante cuidado ao interpretar as citações. É esta parte da história que é mais irônica. No meu caso, começou em 2005 como algo um tanto divertido. Meu amigo e colega Stephen Pratt me perguntou durante uma

pausa para um café: “Você já ouviu falar do índice h ?” Não, eu não tinha ouvido. Ele me explicou: “Para se ter um índice h valendo h você precisará ter publicado h artigos, cada um tendo sido citado pelo menos h vezes.”

Hã? Levou um tempo até minha cabeça se acostumar com isso, então Stephen me mostrou os meus artigos listados no Google Acadêmico. Eu havia publicado apenas nove, na época, um que havia sido citado sete vezes, outro, quatro vezes, e dois outros haviam sido citados três vezes cada. Assim, meu índice h era um mero três. Eu tinha três artigos que haviam sido citados mais de três vezes. Stephen estava na minha frente, com um índice h de seis. Assim que entendi a ideia, olhamos para todos os que conhecíamos. O chefe de Stephen, o eminente ecologista e matemático Simon Levin, tinha um índice h de mais de 100: mais de 100 artigos citados mais de 100 vezes.

Não demorou muito para que todos na área acadêmica estivessem falando sobre citações e o índice h , e não apenas durante as pausas para o café. Políticos e agências de fomento abraçaram rapidamente a ideia de usar as citações para avaliar os cientistas. Finalmente eles possuíam uma maneira de medir o que estava acontecendo no interior das universidades. Por muito tempo, a área acadêmica tinha sido um mundo fechado, no qual os contribuintes tinham que confiar que fôssemos produzir boas ideias. Agora, os políticos e administradores achavam que poderiam usar as citações para mensurar quantas boas ideias produzíamos.

Mais ou menos na mesma época em que Stephen e eu estávamos calculando os índices h um do outro, ouvi o então assessor científico chefe do governo do Reino Unido, Lord Robert May, apresentar um artigo sobre “A riqueza científica das nações”.⁵ Quando ele assumiu seu papel de assessor, Lord Bob quis determinar o quanto o Reino Unido se comparava com outros países, em termos de performance científica. Inicialmente, ele calculou quantas vezes artigos do Reino Unido haviam sido citados e a quantidade de dinheiro gasta na pesquisa. Ele então dividiu o primeiro número pelo segundo e mostrou que o Reino Unido gerou 168 citações por milhão de libras gastas. Era a melhor pesquisa por libra no mundo. Os Estados Unidos e o Canadá estavam em segundo lugar, com 148 e 121,

enquanto o Japão, a Alemanha e a França tinham, todos, menos de 50 citações por milhão de libras gastas. A ciência no Reino Unido se encontrava na posição de líder mundial, por uma margem significativa.

Esta conclusão específica foi amplamente esquecida. Em vez disso, a principal mensagem que o governo do Reino Unido e, nos anos seguintes, os governos de outros países entenderam a partir do artigo de Lord Bob era de que agora poderíamos avaliar de forma confiável o sucesso científico. O resultado foi um amplo monitoramento por algoritmo do estado dos departamentos das universidades. Foi pedido que os acadêmicos enviassem uma lista dos seus artigos recentes como parte de um exercício de avaliação de pesquisa. Por se tratar de artigos novos, eles não poderiam ser avaliados somente com base no número de citações: os artigos não tinham tido tempo ainda de acumular citações. Em vez disso, determinava-se a qualidade do artigo pelo “impacto” da revista científica em que havia sido publicado, e o impacto era avaliado pelo número de citações de outros artigos nessa revista.

A caça pelo impacto levou, por sua vez, a um efeito “também gostaram” em revistas científicas. Os jornais com altos fatores de impacto, i.e., que publicavam artigos que recebiam muitas citações, atraíram mais e mais submissões de qualidade do que aqueles com menor fator de impacto. Jovens cientistas se viram competindo uns com os outros para ter seus artigos publicados nesse pequeno conjunto de revistas com alto prestígio. Em vez de se concentrarem apenas em fazer trabalhos com qualidade, os cientistas estavam fazendo esquemas para encontrar maneiras de elevar seu índice h e assim conseguir publicar seus artigos em revistas com alto fator de impacto.

O efeito “também gostaram” se aplica aos próprios cientistas. Um estudo mostrou que artigos escritos por autores que já haviam feito vários artigos bem citados ganhavam citações mais rapidamente. Não se tratava apenas do número de citações que um artigo tinha, mas a reputação do autor que produzia sucesso.⁶ Monitorar seus próprios registros de citações, e os de seus

colegas, não era mais algo divertido, como havia sido para Stephen e eu — era uma necessidade para sobrevivência acadêmica.

Os cientistas são um grupo de pessoas bastante inteligentes. Quando recebem incentivos para produzir pesquisa de alto impacto, é exatamente isso que eles fazem. Para descobrir a estratégia de publicação ideal, Andrew Higginson, da Universidade Exeter, e Marcus Munafò, da Universidade de Bristol, elaboraram uma analogia matemática entre sobrevivência científica e sobrevivência de animais na natureza por seleção natural.⁷ No modelo de Andrew e Marcus, os cientistas podem tanto investir seu tempo explorando novas ideias ou usá-lo para confirmar resultados de estudos anteriores. Andrew e Marcus mostraram que o ambiente de pesquisa atual, beneficiando o alto impacto, encoraja os cientistas a investirem a maior parte de seu tempo explorando novas ideias. Os cientistas que fazem vários estudos investigando possibilidades originais sobrevivem, enquanto aqueles que ficam verificando cuidadosamente seus resultados “desvanecem”.

À primeira vista, isso pode parecer uma coisa boa: cientistas concentrando suas pesquisas em ideias originais em vez de ficar remoendo os mesmos resultados antigos e sem graça. Mas o problema é que mesmo os melhores cientistas cometem erros sem querer. Esses erros incluem vários resultados que são falsos positivos estatísticos: se muitos experimentos são realizados motivados pela procura por originalidade, então, ao acaso, alguns desses experimentos vão parecer que produziram um resultado novo excitante, enquanto, na verdade, os cientistas “tiveram sorte”.

Essa sorte está com os pesquisadores que obtiveram o resultado. Agora eles podem publicar seus trabalhos numa revista de prestígio e, talvez, conseguir um novo subsídio para pesquisa. Para a ciência como um todo, essas descobertas sortudas, mas incorretas, estão longe de ser positivas. No modelo de Andrew e Marcus, há pouco incentivo para outros pesquisadores verificarem esses resultados. Os cientistas que estavam interessados em confirmar ou refutar a pesquisa de outras pessoas se viram sem trabalho depois de escreverem artigos pouco citados. Como resultado, mais e mais ideias incorretas se acumulam no cânone científico.

Como todos os modelos, o trabalho de Andrew e Marcus é uma caricatura de atividade científica.

Apesar dos problemas, eu não acho que o monitoramento e a caça de citações tenha danificado a qualidade da pesquisa atual realizada pelos cientistas. Isto é, quando temos tempo para fazer pesquisa, nós, cientistas, ainda a realizamos muito bem. A maior parte dos cientistas que encontro são motivados pela busca eterna da verdade e pelo desejo de saber a resposta certa. Nós nos divertimos procurando maneiras de provar que nossos colegas erraram reavaliando seus resultados. Para a maioria de nós, contradizer a teoria de um colega é quase tão gratificante quanto obtermos resultados novos nós mesmos. Assim, os incentivos continuam sendo testar teorias potencialmente exageradas e resultados experimentais incorretos.

O que a avaliação algorítmica de pesquisa fez foi reduzir a quantidade de tempo que temos para nos dedicar à pesquisa pura. Os artigos tiveram que se “tornar sexy”, como dizemos frequentemente, a fim de serem aceitos em revistas importantes. Isso envolve exaltar os melhores resultados e mostrar por que pesquisadores de outras áreas ficarão interessados nesses resultados. Nosso trabalho também é “quebrado”, outra expressão que usamos, em partes menores e publicáveis a fim de maximizarmos o número de artigos que escrevemos. Todo esse tornar sexy e quebrar leva tempo. Temos que escrever mais palavras e submeter e voltar a submeter nossos artigos mais vezes, começando pelas revistas de alto impacto, que rejeitam a maioria dos artigos, até as de baixo impacto.

É aqui que mora a ironia. O algoritmo acadêmico de Anurag nos forneceu um ganho de eficiência de 10%. O que o estabelecimento científico e as entidades financeiras fazem com essa eficiência? Eles a usam para nos monitorar e controlar. Eles nos fizeram mudar a maneira como trabalhamos de modo que perdemos a maior parte da eficiência que ganhamos. E eles criaram uma comunidade científica do tipo “rico fica mais rico”, dominada pelo 1% do topo. Isso agrada muitos cientistas, mas deixa outros cientistas muito bons excluídos.

Nem todos os cientistas se renderam ao concurso de popularidade. Alguns contra-atacaram da maneira que fazem melhor: usando a ciência. Santo Fortunato mostrou que, a longo prazo, o índice h representa apenas um indicador muito fraco de produtividade científica, especialmente quando é usado para avaliar cientistas mais jovens.⁸ Dos 25 pesquisadores que ganharam o prêmio Nobel entre 2005 e 2015, 14 deles tinham o índice h menor que 10 quando tinham 35 anos de idade. Um h de 12 é frequentemente citado como uma exigência para ser contratado numa posição com contratação condicional,⁹ uma regra que teria implicado que esses laureados com o prêmio Nobel não conseguiriam um emprego.

Albert-László Barabási — que, no início de sua carreira, ascendeu a um nível de celebridade acadêmica comparada à de um youtuber com um artigo sobre “também gostando” e gráficos log-log — mostrou que o artigo mais importante que um cientista escreve tem probabilidade igual de surgir a qualquer momento em sua carreira.¹⁰ Pode ser o primeiro a ser escrito. Pode ser um artigo escrito logo após sua defesa de doutorado ou quando ele está dando um duro danado para conseguir uma posição permanente. Pode ser escrito quando ele já tiver se estabilizado como pesquisador ou pode ser o último artigo que ele escrever. A grande descoberta pode acontecer a qualquer momento. Essa observação torna muito difícil para as agências de fomento saberem para quem dar verba, mas sugere que somente citações anteriores não são a solução. Financiar pesquisadores bem-sucedidos pode significar perda de grandes oportunidades, enquanto negligenciar pesquisadores que trabalham durante anos sem fazer uma grande descoberta pode nos privar das nossas descobertas mais importantes.

Algoritmos como “também gostaram” fornecem novas formas de comportamento coletivo, novas maneiras de interagirmos uns com os outros. Isso pode ter muitos efeitos positivos, permitindo que compartilhem nosso trabalho de maneira mais rápida e mais ampla. Mas não devíamos deixar o algoritmo ditar como vemos o mundo. Isso aconteceu em certo grau na academia. Por serem fáceis de se determinar, os índices e impactos de citações tornaram-se a moeda atual na ciência.

A desigualdade é um dos maiores desafios da sociedade, e ela é exacerbada pela nossa vida on-line. Julgamos uns aos outros pelo número de amigos que possuímos no Facebook, pelo número de seguidores que temos no Twitter e pelo número de contatos que temos no LinkedIn. Esses julgamentos não estão completamente errados: como Duncan Watts e Matthew Salganik observaram em seu estudo de paradas musicais no capítulo anterior, músicas realmente ruins ficaram por último. Mas também não estão completamente certos. A mesma crítica que me fez supor que Jake Paul possui apenas um talento módico pode ser aplicada a cientistas bem-sucedidos. Acumular capital social, na forma de curtidas e compartilhamentos, leva ao acúmulo de capital financeiro, seja na forma de bolsas de pesquisa ou Lamborghinis. A retroalimentação então continua.

O algoritmo “também gostaram” é bem simples de ser entendido e, se você não o entendeu direito da primeira vez, volte ao começo do capítulo 9 e releia a descrição do algoritmo. Você precisa entender como ele distorce dados de entrada e de saída, porque está definitivamente operando em algum aspecto de sua vida. Esteja você aumentando sua rede de contatos no LinkedIn para impressionar empregadores ou seja você um empregador que está comparando um candidato impetuoso com uma rede social ampla a um candidato mais discreto com poucos amigos no Facebook, não deveria deixar o algoritmo tomar decisões por você. Nossa integridade humana é uma das coisas mais importantes que temos.

Existem outras maneiras de organizarmos nossa vida on-line.

O algoritmo “também gostaram” não é a única forma de compartilhar informação. Perguntei a Kristina Lerman qual serviço de “compartilhamento” ela achava que tinha o melhor sistema. Ela optou pela versão original do Twitter. Antes de 2016, o Twitter simplesmente mostrava as coisas que eram compartilhadas por pessoas que você seguia em ordem cronológica. Se você não estivesse no site quando um amigo seu publicasse alguma coisa, você provavelmente iria perdê-la. Retweets de outros usuários poderiam ajudar, mas o tempo era o fator principal que determinava o que você via.

Gradualmente, o Twitter migrou na direção do algoritmo “também gostaram”. Ele tem agora o recurso “caso você tenha perdido”, que promove o conteúdo muito curtido e com muitos retweets para o topo de sua linha do tempo. A opção padrão nas configurações é agora “mostrar os melhores tweets primeiro”. Desligue isso. Exponha-se a uma variedade de opiniões mais ampla possível.

Um aplicativo que tem filtragem mínima é o Tinder. Eu não tenho o Tinder e, embora esteja preparado para dissecar a maior parte da minha vida on-line a fim de entender como os algoritmos funcionam, estabeleci um limite com relação a baixar um app de encontro on-line. Minha mulher, compreensiva como ela é, não iria me perdoar por criar uma conta, mesmo que eu argumentasse que é para fins científicos.

Vários de meus colegas mais jovens ficaram entusiasmados em me explicar o Tinder, uma vez que este ocupa grande parcela de seu tempo. Sendo um usuário, você verá uma série de fotos de perfil de pessoas em quem possa estar interessado e que estejam perto de você. Deslize para a direita se gostar do que vê, deslize para a esquerda se não estiver interessado. Se você deslizar uma pessoa para a direita e esta pessoa deslizar você para a direita, então poderão conversar por intermédio do app e, com sorte, um romance (ou seja lá o que você estiver procurando) irá florescer.

Com a ênfase na foto, o perfil do usuário contém apenas uma biografia curta, incluindo seu nome, idade, interesses e um pequeno texto sobre si mesmo. Quando lançado, o Tinder foi um antídoto para outros sites de encontro on-line que se vangloriavam de seus algoritmos complexos projetados para encontrar o par perfeito. Cansados de preencher questionários ou de ter seu perfil do Facebook analisado, os jovens deram boas-vindas à simplicidade e à honestidade. Sem filtros, o Tinder lhe permite avaliar o que está disponível para você.

Existe uma diferença enorme em como homens e mulheres deslizam. Em um estudo conduzido em Londres usando contas falsas e uma única foto de perfil, as mulheres eram 1.000 vezes mais propensas a receber uma deslizada para a direita de um homem do que os homens a receber de uma mulher.¹¹

Se você é um desses homens solitários que não têm recebido nenhuma deslizada para a direita, há algumas coisas que pode fazer. Primeiramente, escrever uma biografia quadruplica a probabilidade de ser escolhido, assim como incluir duas fotos de perfil adicionais. As mulheres querem mais informação e são mais seletivas do que os homens, assim, se você quer um *match*, precisa fazer o que elas querem.

Essas dicas básicas não resolvem completamente os problemas que os homens estão tendo. Gareth Tyson, que conduziu o estudo de Londres, também enviou um questionário para descobrir quantos *matches* (deslizadas para a direita por ambas as partes) as pessoas conseguiam. A maioria dos homens conseguiu *matches* em menos de 10% de suas deslizadas para a direita. Um grupo sortudo de 5% conseguiu mais de 50% de *matches*. Apesar de o algoritmo ser muito diferente, a desigualdade é muito parecida com aquela vista entre artigos científicos: um grupo bem pequeno de homens provavelmente muito atraentes ficou com toda a atenção. O restante dos homens tem que deslizar muito para conseguir um *match*.

Será?

Alex, meu colega com quem analisei o FiveThirtyEight, tem muita experiência pessoal com o Tinder. Ele foi para a Suécia como um solteiro australiano e, a princípio, o aplicativo de encontros românticos pareceu a forma ideal para conhecer pessoas novas. Mas ele logo percebeu que era um dos homens que não estava conseguindo encontros e se perguntou o que poderia estar fazendo de errado. Sua resposta foi criar um modelo matemático de como homens e mulheres se comportam no site. Quando você não consegue marcar encontros, sobra mais tempo para a matemática, eu acho.

Alex percebeu que havia uma conexão entre não conseguir um *match* e deslizar mais. Quando os homens usam o Tinder pela primeira vez, eles tendem a ser muito criteriosos, mas ao perceberem que não estão conseguindo *matches*, eles ampliam as buscas e deslizam para a direita ainda mais. As mulheres fazem o oposto. Como estão recebendo *matches* demais, elas reduzem a busca. O modelo de Alex mostrou que essa conexão tornava

a situação ainda pior para todo mundo. No final, mulheres escolhiam apenas um ou dois homens e os homens escolhiam praticamente todas as mulheres. Alex chamou isso de “Jogo de encontros instável”, porque a solução estável para encontrar os pares perfeitos desapareceu para homens e mulheres.¹² Em vez disso, o resultado do algoritmo foi o mesmo que Gareth Tyson observou em seu estudo dos usuários de Londres: na média, homens deslizavam para a direita cerca de metade das mulheres que viam no aplicativo, enquanto a maioria das mulheres deslizava para a direita menos de 10% dos homens.

Alex encontrou uma solução que funcionou para ele. Ele foi na direção oposta do rebanho. Ele decidiu se tornar mais paciente e mais exigente nas suas escolhas de encontros. Ao buscar com cuidado mulheres em quem ele realmente acreditava, escrevendo uma biografia que imaginava que elas fossem achar interessante e, então, esperando até que algumas delas o escolhessem também, sua taxa de *matches* aumentou drasticamente. Ele ainda não encontrou seu amor verdadeiro no Tinder, mas já teve vários encontros agradáveis em cafés de Estocolmo, e até mesmo começou uma banda com uma das mulheres com quem saiu.

Um sistema “também gostaram” para encontros não funcionaria de fato. “Sim, ele é ótimo. Talvez você deva fazer um teste com ele?”, não é uma noção que normalmente escutamos ser compartilhada entre amigas, on-line ou não. Mas um sistema de deslizamentos para artigos acadêmicos é uma ideia em que vale a pena investir. A cada manhã, quando chego ao escritório, eu poderia me sentar e deslizar os artigos de meus colegas para a esquerda e para a direita sem um algoritmo me dizendo o que meus pares citaram. Se deslizássemos corretamente uns aos outros, este aplicativo poderia nos colocar em contato (diretamente sem ninguém saber que somos amigos) com colegas para que pudéssemos conversar um pouco sobre ciência.

Anurag Acharya, se você estiver lendo, talvez possa encarar essa. E, finalmente... poderei deslizar sem acabar com meu casamento.

Nota

* A proporção de artigos que são citados p vezes ou mais é proporcional ao número de n vezes que um artigo é citado, elevado a uma potência numérica constante $-a$ (i.e., $p = kn^{-a}$, onde k é também uma constante⁴).

BORBULHANDO

A votação a favor do Brexit e a eleição de Donald Trump foram grandes surpresas para aqueles de nós que vivem protegidos pela academia. A maioria dos meus colegas liberais não entendeu o que estava acontecendo. Eles não conheciam ninguém que votaria em Trump, e nunca conheceram alguém que quisesse sair da União Europeia. Os jornais liberais que eles liam estavam igualmente surpresos, publicando matérias com títulos do tipo: “Conheça dez britânicos que votaram pela saída da UE” e “Por que a classe branca trabalhadora votou em Trump”. Havia uma necessidade desesperada para explicar como e por que os eleitores, de uma hora para outra, não concordavam com o que a maioria de nós considerava ser um consenso estabelecido sobre como as coisas devem ser feitas.

Eu estava tentando interpretar o imenso número de artigos que haviam sido escritos sobre a política do ano anterior. Uma explicação recorrente era que os algoritmos seriam os culpados por informar as pessoas erroneamente. O *New York Times*, o *Washington Post*, o *Guardian* e o *Economist* estavam entre os muitos meios de comunicação nos quais foram publicadas reportagens, que soavam como matemática, sobre o isolamento e a polarização criados por algoritmos.

Para começar, a discussão era sobre câmaras de eco e bolhas-filtro. A teoria era de que o Facebook e o Google estavam personalizando nossas

buscas de tal modo que só víamos o que queríamos ver. O foco, então, se voltou para as fake news. Adolescentes da Macedônia estavam gerando automaticamente notícias novas, testando diferentes combinações de rumores sem sentido sobre Trump e Clinton, a fim de gerar tráfego em seus sites e faturamento com propaganda. Foi dito que a Rússia estava influenciando a eleição usando anúncios no Facebook e trolls da internet pagos, que discutiriam agressivamente no Twitter ou em vários blogs políticos a fim de criar polarizações.

Muitas das preocupações sobre matemática e algoritmos ressoaram com o que eu já havia descoberto sobre o “também gostaram” e os algoritmos que o Facebook usava para analisar personalidades. Parecia que estávamos deixando os algoritmos nos dizer o que pensar e fazer. Havia um risco genuíno de que algumas das notícias que nos foram mostradas serem distorcidas ou inventadas por chapéus pretos politizados. Mas me senti desconfortável com a forma como a matemática estava sendo usada nessas reportagens e o que elas insinuavam sobre as pessoas que estavam absorvendo a mídia. Eram os americanos assim tão estúpidos a ponto de as únicas mensagens que eles não filtraram serem aquelas geradas por adolescentes da Macedônia e trolls russos? As pessoas estavam assim tão influenciadas pelo que viam no Facebook? Muitos dos meus colegas achavam que sim. Eu não estava tão certo.

Muito antes de os acadêmicos e jornalistas estarem se preocupando com as bolhas-filtro de Trump e de Clinton, dois jovens cientistas da computação já estavam pesquisando como as campanhas políticas influenciavam e eram influenciadas pela internet. Em 2004, Lada Adamic e Natalie Glance estudaram a “blogosfera” na campanha presidencial daquele ano.

Em comparação com as redes sociais atuais, esses blogs parecem ser um tanto pitorescos. O formato era simples: textos explicando o ponto de vista do blogueiro, algumas imagens de jornais coladas e links para sites de notícias e outros blogs. Não havia botões “curtir” e “compartilhar” conectados às redes sociais. O Facebook ainda não havia decolado, e o Twitter estava para ser inventado. Em vez disso, blogs políticos se

conectavam por intermédio de links diretos entre páginas da internet, muitas vezes na forma de “lista de blogs”, listando sites e artigos aprovados pelo autor do blog.

A Figura 11.1 mostra os links entre os 20 maiores blogs liberais (simpatizantes do partido Democrata, círculos pretos) e os 20 maiores blogs conservadores (simpatizantes do partido Republicano, círculos cinza) na campanha eleitoral de 2004. O tamanho do círculo representando cada blog é proporcional à sua popularidade entre os outros blogs. A espessura da linha indica o número de links entre os blogs.

A blogosfera de 2004 estava dividida. Os blogs democratas se conectavam quase que exclusivamente a blogs democratas. Blogs republicanos se conectavam quase que exclusivamente a blogs republicanos. Havia muito poucos links entre os dois mundos. O único blog republicano com mais do que apenas alguns poucos links para um blog democrata era o The Daily Dish, administrado por Andrew Sullivan (representado por AS na Figura 11.1), e ele havia trocado sua aliança política na campanha de 2004 para apoiar o democrata John Kerry. Fora isso, a segregação era em sua maior parte completa.

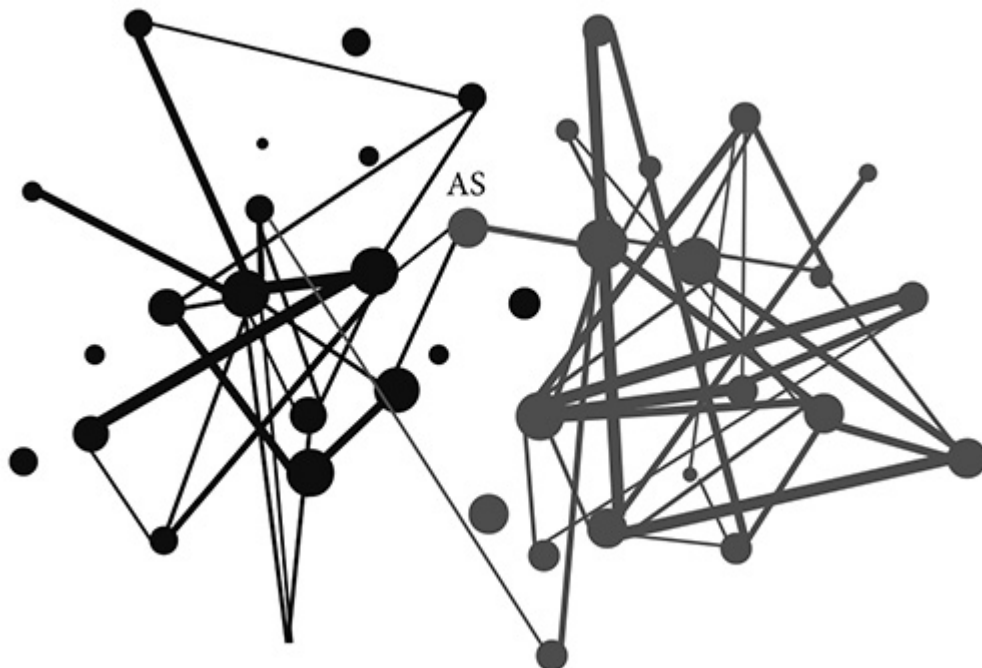


Figura 11.1 Rede de conexões entre os 40 maiores blogs políticos na campanha presidencial de 2004. Círculos pretos indicam blogs liberais. Círculos cinza indicam blogs conservadores. A área dos círculos é proporcional ao número de vezes que um blog é conectado a outro dos 40 maiores blogs. A espessura das linhas é proporcional ao número de vezes que um dos blogs é conectado a outro, em que são mostrados somente blogs com cinco ou mais links. O blog rotulado por AS é o The Daily Dish, administrado por Andrew Sullivan. Elaborada com base nos dados do artigo: “The political blogosphere and the 2004 US election: divided they blog”, por Adamic and Glance (2005).

As redes de blogs democratas e republicanas eram diferentes umas das outras. Se você analisar cuidadosamente a Figura 11.1, verá que blogs conservadores têm mais conexões do que blogs liberais. Blogueiros conservadores tendiam a fazer mais comentários nos textos dos outros do que os liberais. Houve mais discussão interna ativa entre os simpatizantes republicanos. O mais importante, contudo, é que nem a rede liberal nem a conservadora estavam fechadas ao mundo exterior. Praticamente uma publicação sim, outra não, de ambas as convicções políticas, continham uma referência a uma matéria jornalística na mídia convencional. Por exemplo, o *Washington Post* foi citado 900 vezes por blogs liberais e 500 vezes por conservadores durante os dois meses e meio que antecederam a eleição. Os blogs liberais e conservadores estavam segregados, mas ambos abriam espaço para a mídia convencional.

O estudo de Lada e Natalie era uma dica de como os cientistas da computação analisariam notícias e política no futuro. Os métodos que elas desenvolveram durante seu estudo — para dissecar a rede política, identificando automaticamente as palavras-chaves nos debates e descobrindo como comentadores de política se conectavam — estavam mudando o modo como entendíamos a mídia. No artigo, elas escreveram que, no futuro, “vamos querer monitorar a propagação de notícias e ideias nas comunidades e identificar se os padrões de conexão na rede afetam a velocidade e o alcance da propagação”.

Em um contexto acadêmico, havia, portanto, muitas possibilidades excitantes. O estudo de Lada e Natalie mostrou que era possível usar métodos matemáticos para entender como nos comunicávamos sobre

política. Mas se cientistas conseguem desenvolver métodos para entender comunicação política, então os partidos políticos e as grandes empresas podem usar esses métodos para manipular como *nós* conversamos uns com os outros. E, a partir de 2004, esses métodos se desenvolveram muito rapidamente.

Na eleição de 2016 dos Estados Unidos e na votação do Brexit, Breitbart e Huffington Post haviam se tornado o amálgama de direita e de esquerda desses blogs políticos iniciais. Muitos dos blogueiros de 2004 estavam trabalhando para esses sites, ou outros, como o Drudge Report. Alguns, como Andrew Sullivan, estavam escrevendo artigos sobre como eles ficaram cansados de suas atividades constantes em redes sociais. Mas havia muitas vozes novas para substituí-los: blogs políticos independentes estavam aparecendo do nada o tempo todo, e dezenas de milhares de pessoas já usavam a plataforma de publicação on-line Medium para documentar cada um de seus pensamentos. Artigos são “votados positivamente” e “votados negativamente” no Reddit; eles são espalhados em sites de tendência como o BuzzFeed e Business Insider; são agregados no Feedly e no Fark; transformados em jornais espontâneos no Flipboard; e compartilhados no Facebook. Todos esses milhões de palavras, escritas todo dia de todos os cantos do planeta, são então comentadas, discutidas, ridicularizadas e/ou apoiadas usando os 280 caracteres fornecidos pelo Twitter.

Quando comentaristas analisam essa vasta gama de mídia social, normalmente eles retornam a dois temas principais: a câmara de eco e a bolha-filtro. Esses conceitos estão relacionados, mas são ligeiramente diferentes. A rede de blogs políticos de 2004 é um exemplo primitivo de câmara de eco. Blogueiros conectados a outros blogueiros que concordam com eles, confirmando seus pontos de vista e dando suporte ao que eles já pensavam. Se você clicasse de um blog para outro, escolhendo links aleatórios de cada página que visitasse, continuaria preso dentro do mesmo conjunto de opiniões com o qual começou. Se você começasse numa página de um blog liberal em 2004, então a probabilidade de que 20 cliques depois você ainda estivesse numa página liberal seria de 99%. Comece numa página

conservadora e, então, 20 cliques depois você estará provavelmente lendo material conservador. Cada conjunto de blogueiros havia criado seu próprio mundo, dentro do qual suas opiniões reverberam.

As bolhas-filtro vieram depois e ainda estão em desenvolvimento. A diferença entre cavidade “filtrada” e “ecoada” está no fato de terem sido criadas por um algoritmo ou por pessoas. Enquanto os blogueiros escolhem links para blogs diferentes, os algoritmos baseados em nossas curtidas, buscas e histórico de navegação não envolvem uma escolha ativa de nossa parte. São esses algoritmos que podem, potencialmente, criar uma bolha-filtro.¹ Cada ação que você faz no seu navegador é usada para decidir o que lhe será mostrado depois.

Sempre que você compartilha um artigo, por exemplo, do jornal *Guardian*, o Facebook atualiza seu banco de dados para refletir o fato de que você está interessado no *Guardian*. De forma análoga, quando você compartilha algo do *Telegraph* o banco de dados armazena um interesse por esse jornal. Em abril de 2016, numa palestra para editores, Adam Mosseri, responsável pelo desenvolvimento do algoritmo do feed de notícias do Facebook, explicou que “quando você se registra no Facebook, seu feed de notícias é uma página em branco. Lenta, mas seguramente, você adiciona pessoas de quem gosta e segue usuários que lhe interessam. E você constrói sua própria experiência personalizada”.²

O algoritmo do Facebook decide que informação nos vai ser mostrada com base nas escolhas que já fizemos. Para entender como seu filtro funciona, suponha que você seja uma pessoa relativamente de mente aberta e independente que acabou de se registrar no Facebook. Vamos supor que tenha a mente tão aberta que lê tanto o jornal de direita *Telegraph* quanto o jornal de esquerda *Guardian*, e que compartilha artigos das duas publicações. Vamos supor também que você tenha amigos de direita e de esquerda em quantidades aproximadamente iguais. Eu sei que isso é muito difícil de acontecer na vida real, mas a fim de perceber o possível efeito do algoritmo, começamos supondo que você seja um indivíduo com a mente mais aberta possível.

Agora, você começa a publicar. Suponha que você faça algumas poucas publicações compartilhando artigos do *Guardian* e do *Telegraph*. Seus amigos não percebem no início, mas depois um deles faz um comentário em um link que você compartilhou do *Telegraph* sobre problemas relativos à corrupção dentro da União Europeia. Você responde, e vocês discutem sobre o artigo, curtindo as publicações um do outro. Agora você está dando ao Facebook o que ele quer: informação sobre o que faz você passar um tempo no site. Em troca, o Facebook pode lhe dar mais do que ele acha que você deseja. No dia seguinte, uma publicação de seu amigo, lamentando-se sobre a nova regulamentação da UE, aparece no topo de seu feed de notícias. O artigo logo abaixo é do *Telegraph*; outro amigo compartilhou informações sobre vantagens de negócios para um Reino Unido pós-Brexit. Ambos os artigos chamam sua atenção e você começa a comentar. O Facebook percebe seu interesse adicional e no dia seguinte fornece outros artigos sobre críticas à UE. O filtro vai se formando lentamente ao seu redor.

Essa descrição é simplesmente uma única história, mas podemos usar um pouco de matemática para entender como o algoritmo de filtro vai funcionar tipicamente. O Facebook determina o quão visível um artigo de jornal compartilhado recentemente ficará em seu feed de notícias baseado parcialmente na seguinte equação:³

$$\text{Visibilidade} = (\text{seu interesse pelo jornal}) \times (\text{proximidade ao amigo que compartilha o artigo})$$

Quando interagiu com seu amigo sobre a publicação que você compartilhou, você acionou ambos os fatores dessa equação: mostrou maior interesse no *Telegraph* e fez o Facebook aumentar o nível de proximidade entre você e seu amigo. Desse modo, podemos ver o aumento da visibilidade como um engajamento ao quadrado, i.e., a visibilidade é proporcional ao engajamento por meio do interesse no jornal multiplicado pelo engajamento por intermédio da proximidade ao amigo. Como resultado, a visibilidade de

artigos futuros do *Telegraph* aumenta. Uma visibilidade aumentada faz com que seja mais provável que você clique nesses links no futuro, o que aumenta ainda mais a classificação do Facebook e leva a ainda mais visibilidade.

O próximo passo é colocar tudo isso num modelo matemático que capture tanto o comportamento do algoritmo quanto a interação do usuário com o algoritmo. Vou descrever isso, agora, na forma do que eu chamo de modelo “filtro”. Assim como no meu modelo “também gostaram” da Amazon, o modelo “filtro” é uma simplificação de como o algoritmo do Facebook funciona de fato. Ele captura o recurso mais central por meio do qual o Facebook filtra nosso feed, o Twitter filtra nossa linha do tempo e o Google filtra nossas buscas: quanto mais clicamos em alguma coisa, ou alguém, mais proeminentemente eles nos são mostrados, e mais provavelmente iremos continuar a clicar neles.

O modelo “filtro” funciona em termos de um número de interações. A cada interação, são mostradas a um usuário publicações de duas fontes; no meu exemplo anterior, essas fontes são o *Guardian* e o *Telegraph*. Supondo que a visibilidade desses dois jornais seja determinada pela equação de visibilidade dada anteriormente, e que os usuários clicam num jornal proporcionalmente à sua visibilidade, podemos simular como a visibilidade relativa desses dois jornais varia com o tempo.⁴

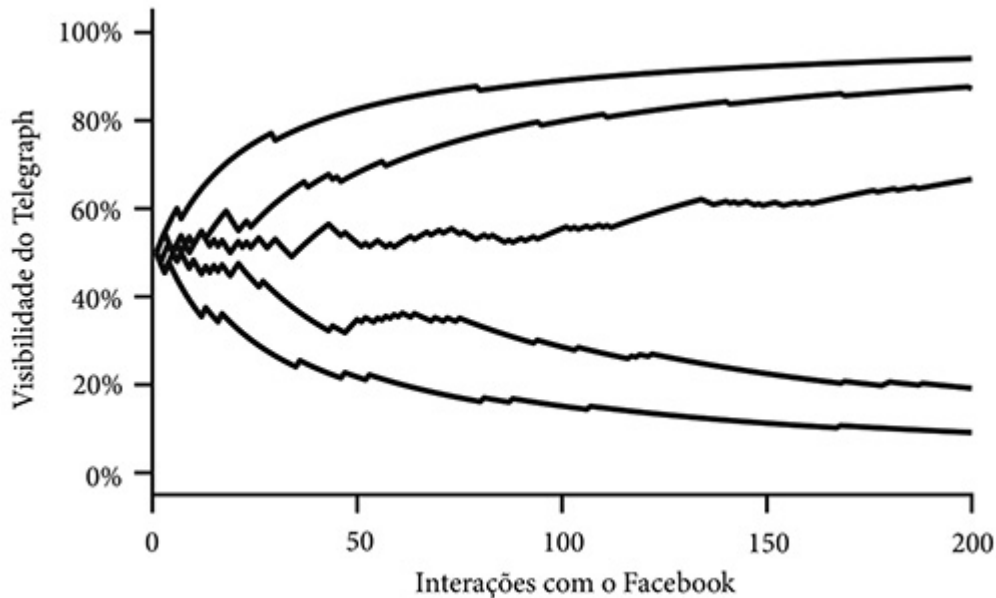


Figura 11.2 Simulação de como o modelo “filtro” funciona para cinco usuários “imparciais” diferentes. Em cada rodada do modelo, o usuário escolhe se vai interagir com uma publicação do Guardian ou do Telegraph. Se eles escolhem o Guardian, então a visibilidade dos artigos do Guardian aumenta, assim como a probabilidade de, na próxima rodada, eles escolherem o Guardian novamente. Esse feedback leva, no final das contas, os usuários a aumentar a visibilidade de um dos jornais em vez do outro.

A Figura 11.2 mostra como a visibilidade de artigos do *Guardian* e do *Telegraph* muda para cinco usuários simulados diferentes, os quais começaram com 50% de visibilidade para ambos os jornais. Depois de 200 interações com o Facebook, dois dos usuários simulados no algoritmo “filtro” têm um alto nível de visibilidade para o jornal *Guardian*, dois têm um alto nível de visibilidade para o jornal *Telegraph*, e o último tem uma visibilidade ligeiramente maior para o *Telegraph*. A simulação com mais usuários produz resultados parecidos: depois de 200 rodadas do filtro, são mostradas para a maioria dos usuários publicações de um dos dois jornais.

Lembre-se de que esses usuários simulados não tinham uma preferência inicial por nenhum dos jornais. A preferência se forma à medida que os cliques mudam a visibilidade. O feedback entre as escolhas que eles fizeram e as publicações que lhes são mostradas determina qual jornal lhes será mostrado mais frequentemente.

Como Adam Mosseri diz, começar com “você é uma página em branco”. Mas assim que os usuários escrevem sua primeira mensagem naquela página, o feedback entre a visibilidade e o engajamento começa. Uma vez que a visibilidade é proporcional ao engajamento multiplicado por ele mesmo, a página é rapidamente preenchida com o jornal pelo qual, por acaso, o usuário mostrou interesse primeiro. O algoritmo “filtro” cria uma bolha, mesmo para pessoas inicialmente imparciais.

Para pessoas que começaram a usar o Facebook já com uma preferência preestabelecida com relação à mídia de direita ou de esquerda, o efeito é ainda mais acentuado. O algoritmo “filtro” percebe pequenas diferenças iniciais e as exagera até que o lado oposto do argumento seja perdido. Os usuários ficam presos em ideias autocorroboradas e interações com um grupo menor de amigos.

O algoritmo que o Facebook usa em seu feed é um pouco mais complicado que meu modelo “filtro”. Ele alega usar mais de 100.000 fatores personalizados para tomar decisões sobre o que lhe mostrar (o que quer dizer, na verdade, que ele fez uma análise de componente principal de suas “curtidas”). Assim, embora o modelo mostre que existe um risco de que o Facebook crie bolhas-filtro, ele não prova que todos os usuários estejam aprisionados em bolhas. Eu queria descobrir o quanto o modelo do filtro simplificado captura a realidade.

A matemática Michela Del Vicario é uma das pesquisadoras de um grupo de pesquisa do Laboratório de Ciências Sociais Computacionais de Lucca, na Itália, que está colocando o Facebook à prova. Os pesquisadores identificaram 34 páginas do Facebook italiano que compartilham avanços da ciência e 39 que compartilham teorias da conspiração.⁵ Eles estudaram como os usuários do Facebook compartilharam, curtiram e comentaram as publicações feitas nessas páginas. Havia muitas evidências de polarização entre essas duas comunidades. Pessoas que “curtem” e compartilham ciência raramente “curtem” e compartilham conspirações e vice-versa. Também havia evidência de uma câmara de eco dentro de cada comunidade. A maioria das conspirações se espalham principalmente entre as pessoas que

já estão “curtindo” e compartilhando teorias da conspiração e causam pouco impacto na população italiana como um todo.

Quando conversei com Michela, ela descreveu um ciclo vicioso: “Quanto mais artigos sobre conspiração uma pessoa compartilha, maior a probabilidade de ela compartilhar mais dessas histórias e maior a probabilidade de ela interagir com outras pessoas interessadas em teorias da conspiração.” Esse é o processo descrito no meu modelo filtro: existe um feedback entre compartilhar conspirações e o aumento da exposição e compartilhamento de histórias parecidas.

Lamentavelmente, o mesmo acontece com a ciência. Um grupo pequeno de cientistas geeks italianos compartilha as últimas novidades sobre ciência uns com os outros, enquanto a maior parte do público em geral presta pouca ou nenhuma atenção a eles.

Michela e seus colegas também analisaram as palavras usadas nas publicações sobre conspiração e ciência. Ela me disse: “Exceto poucos usuários muito ativos, que são bastante positivos quando publicam, a regra geral é que quanto maior é a atividade do usuário, mais negativas são as palavras que ele usa.”

Embora tanto cientistas quanto teóricos da conspiração tendessem a ser menos positivos conforme mais eles postavam, o efeito era mais intenso entre os cientistas. Os cientistas usaram menos palavras positivas que os teóricos da conspiração, e quanto mais ativos eles eram no Facebook, mais negativos eles ficavam. Tornar-se um membro ativo de uma câmara de eco não é um caminho para a felicidade.

Os teóricos da conspiração não são apenas menos ranzinhas do que os cientistas, mas suas publicações compartilhadas são também mais populares do que aquelas sobre novidades na ciência. Isso é muito preocupante, uma vez que grande parte das teorias da conspiração são *sobre* ciência. Um dos rumores constantes é sobre uma ligação entre vacinas e autismo. Eles persistem apesar de repetidos estudos científicos rigorosos, compartilhados por cientistas, que mostram que não existe tal ligação. Outra teoria em voga e em andamento é a chamada “conspiração dos rastros químicos”, na qual

governos estariam espalhando agentes químicos tóxicos e infecções em nuvens com o auxílio de aviões.

Ao assistir a vídeos sobre rastros químicos, fiquei genuinamente surpreso — não tanto com o conteúdo do filme, mas com minha reação a eles. Sentado sozinho em casa, de frente para minha tela, numa noite tranquila após um dia de trabalho, pude perceber que eu estava lentamente começando a acreditar no que assistia. Um vídeo com 580.842 visualizações no YouTube mostrava um encontro na Califórnia. Depoimentos de pilotos, médicos, engenheiros e cientistas estão reunidos. O cenário é uma audiência governamental local, a sala está lotada e a impressão é a de que uma investigação importante está em andamento. Homens grisalhos se levantam, um depois do outro, e fazem declarações seguras sobre “a natureza penetrante das nanopartículas”, “aumento da poluição do ar” e “redução dramática nas espécies de insetos”. Eles falam sobre doença de Alzheimer, autismo, TDAH, colapso de ecossistema e poluição de rios. Algumas das coisas que eles falam fazem sentido do ponto de vista científico, mais especificamente sobre a qualidade da água e problemas ambientais, e me vejo sendo sugado por seus argumentos. O filme muda com rapidez de um palestrante para o próximo e as câmeras repetidamente enquadram uma plateia que bate palmas apoiando os palestrantes. Eu tenho perguntas a fazer. Mas tudo acontece tão rápido, e não sei o contexto de cada depoimento. Não consigo identificar exatamente o que está errado.

Esse é o efeito de um filme bem produzido sobre conspiração científica. Eu fico perplexo com a justaposição de fato e ficção, tentando entender o amontoado de ideias. Clico num vídeo depois do outro e acabo gastando mais de três horas navegando em vídeos, cada um deles já fora assistido centenas de milhares, ou até milhões, de vezes. Eu assisto a uma apresentação de um membro de semblante austero da Geoengineering Watch, uma organização contra rastros químicos, sobre “a evolução rápida da ciência” por trás dos rastros químicos. Acho um vídeo do Prince falando sobre como a ligação entre os rastros químicos e a violência serviu de inspiração para suas músicas. Há uma mãe preocupada explicando com

calma as conexões entre metais pesados e a saúde humana. Esses filmes são encerrados com confissões de oficiais do governo aposentados e uma mulher que supostamente teve seus filhos afastados dela por ter ousado revelar a verdade sobre como nosso governo nos está envenenando. Após isso, fechei meu laptop e fiquei sentado no escuro. Leva alguns minutos até que minha mente se esvazie e meu cérebro científico comece a trabalhar novamente.

Não acredito que esteja correndo um sério risco de ser sugado para dentro de uma bolha de conspiração, mas assistir a esses vídeos me permitiu entender melhor quem são essas pessoas. Vou em frente e analiso um site sério e factual, organizado por um professor de Harvard de Geoengenharia, David Keith, em que ele explica cuidadosamente por que a conspiração dos rastros químicos é inteiramente fictícia. Eu assisto a um resumo de um vídeo gravado por Steve J. Davis, pesquisador de Ciência do Sistema Terrestre, da Universidade da Califórnia, que entrevistou 77 cientistas sobre a possibilidade de uma conspiração de rastros químicos. Todos os especialistas, com a exceção de um, disseram que as supostas evidências poderiam ser explicadas por outros fatores já bem compreendidos.

Mas o vídeo de Steve só havia sido assistido 1.720 vezes, e eu sabia o motivo. Ele era factual, desprovido de drama. No vídeo, Steve estava casualmente sentado em seu escritório, falando com uma voz neutra sobre a importância da revisão por pares. Sem os interesses de uma organização como a Geoengineering Watch, ele não sentiu necessidade de insistir em seu argumento. Eu consigo entender por que ele apresentou seu trabalho do modo como o fez, mas, ao comparar seu vídeo com os outros a que eu havia assistido anteriormente, pude perceber também por que a bolha científica é muito menor do que a bolha conspiratória. As conspirações são muito mais cativantes.

Dentro de câmaras conspiratórias, as ideias passam incontestadas. As mesmas pessoas ficam compartilhando o mesmo material umas com as outras. Mike Wood, professor da Universidade de Winchester e colaborador do blog A Psicologia das Teorias da Conspiração, me disse que “dentro de

comunidades estabelecidas, vídeos conspiratórios frequentemente têm comentários dissidentes ocultados”. Muitos canais de conspiração do YouTube desativaram a opção de se fazerem comentários, outros estão cheios de mensagens explicando que aqueles que fazem comentários que contradizem a conspiração fazem parte da conspiração. Eles dizem que os anticonspiração são a prova de que a conspiração existe.

Mike gosta de discutir sobre teorias da conspiração na internet. Seu blog tem uma seção de comentários extensa em que ele responde às perguntas com sinceridade, não importando o quanto elas são insensatas ou ignorantes. Apesar do tom paciente de Mike, essa seção de comentários algumas vezes se torna um fórum em que teóricos da conspiração e esclarecedores da anticonspiração se insultam uns aos outros. A seção sobre a conspiração de programação preditiva estava lotada de comentários feitos por dois interlocutores acusando-se um ao outro de “embromação desonesta”, acreditar em “papo furado alienígena”, ter “uma visão paranoica do mundo” e, finalmente, “você fica aí só falando merda”. Mike não fez mais nenhum comentário.

De acordo com Michela Del Vicario, os debates agressivos servem apenas de combustível para os teóricos da conspiração. Ela percebeu que quanto mais comentários negativos um teórico da conspiração encontra, mais provavelmente ele continuará a compartilhar, comentar e discutir sobre eles. Ataques agressivos do exterior só servem para reforçar a bolha.

Existe uma conclusão de ambos os trabalhos de Michela e Mike que oferece um fio de esperança na batalha contra as teorias da conspiração. Assim que teorias da conspiração começam a se espalhar, elas encontram uma resistência maior, tanto por pessoas tipicamente interessadas em ciência quanto por membros do público em geral. Os comentários abaixo dos vídeos mais assistidos sobre rastros químicos continham explicações científicas sucintas (“esse é o vídeo de um avião despejando combustível”), argumentos sensatos de por que a teoria era improvável (“se o governo está realmente o envenenando, por que fazer isso com um rastro opaco de gás quando poderia simplesmente envenená-lo com agentes químicos que você

não consegue ver?”), juntamente com uma série de insultos condescendentes (“espero que você não possa votar nem se multiplicar” e “RETARDADOS!!! Vocês são idiotas”).

Esses comentários garantem que o público em geral, aqueles fora da bolha conspiratória, possa perceber facilmente que uma publicação em particular no Facebook ou um vídeo do YouTube é controverso. Mike me disse: “Se houver apenas um dissidente, ele com frequência será ignorado ou acabará em uma discussão de 50 comentários com uma única pessoa, mas se o vídeo receber atenção de fora, os comentários dissidentes começam a receber curtidas positivas do exterior.” Minha própria pesquisa limitada sobre vídeos conspiratórios — eu já estava na minha quarta noite consecutiva de visualização compulsiva — coincidia com os resultados deles. Eu até passei a “gostar” dos dissidentes e a “não gostar” dos defensores. Segui, porém, o conselho da Michela e resisti em adicionar meus próprios comentários sarcásticos.

A maioria das teorias da conspiração são guardadas dentro de uma bolha, na qual os teóricos apoiam as ideias uns dos outros. Se a bolha se expandir muito, então os anticonspiração podem estourá-la. Câmaras de eco maiores fornecem espaço para mais vozes dissidentes.

As únicas pessoas que têm acesso a dados do Facebook suficientes para dissecar completamente um debate político de larga escala são os cientistas que trabalham lá. Após publicar seu artigo sobre blogosfera política, Lada Adamic se tornou uma pesquisadora de liderança no campo da ciência social computacional. Ela trabalhou como pesquisadora da Ciência da Computação na Universidade de Michigan e publicou alguns dos artigos mais influentes sobre matemática aplicada à ciência social. Então, em 2013, ela aproveitou seu ano sabático para trabalhar no Facebook, onde acabou se estabelecendo como uma cientista de dados.

Trabalhar no Facebook permitiu à Lada testar a hipótese de bolhas-filtro na política dominante. Juntamente com outros dois cientistas do Facebook, Lada analisou a rede de conexões entre amigos no Facebook afiliados politicamente. As redes de amizades no Facebook não eram nem de perto

segregadas politicamente como os blogs políticos de 2004: 20% dos amigos de liberais eram conservadores, 18% eram moderados e 62% eram liberais. Assim, enquanto certamente havia uma tendência tanto pelos liberais quanto pelos conservadores em preferir a amizade de pessoas com a mesma opinião, ambas as convicções políticas estavam expostas a uma boa parte de pessoas que não concordavam com seus pontos de vista.

Os pesquisadores do Facebook analisaram então o conteúdo das notícias compartilhadas pelos amigos dos conservadores e dos liberais. Cerca de 34% do conteúdo compartilhado pelos amigos dos conservadores era composto de notícias liberais. Se compararmos esse número com a quantidade média de notícias liberais compartilhadas no Facebook, que é de 40%, então vemos que a tendência a favor de notícias liberais é pequena. Amigos do Facebook forneceram um leve cantarolar de opiniões, diferentemente de uma câmara repleta de opiniões ecoadas. Amigos de liberais se mostravam mais propensos a compartilhar notícias que refletiam esses pontos de vista. Amigos de liberais compartilharam 23% de conteúdo conservador, comparado à média de 45% em todo o Facebook. Aqui o eco era audível, mas longe de ser um ruído ensurdecedor.

A maioria de nós reconhece essa imagem diversificada dos amigos do Facebook e o que eles compartilham. Temos vários amigos da escola, muitos dos quais nem eram nossos amigos de fato naquela época. Temos amigos do trabalho e de lugares que visitamos nas férias, estudamos e moramos. Provavelmente, já viajei um pouco mais do que um usuário médio do Facebook e agora moro fora do Reino Unido, mas, em muitos aspectos, sou um homem tipicamente inglês. O número médio de amigos do Facebook de um usuário é 200.⁶ Eu tenho 191 amigos. Eles têm origens completamente diferentes e, apesar de serem mais inclinados ao liberalismo de esquerda, possuem uma grande variedade de opiniões políticas.

Sabemos que o Facebook não trata nossos amigos igualmente. A equação de visibilidade/proximidade (veja página 151) implica que os amigos com os quais mais interajo aparecem na minha linha do tempo. Não tenho interagido muito com meus velhos amigos da escola, e o algoritmo do

Facebook sabe disso. Ele não os mostra com tanta frequência na minha linha do tempo. Lada e outros pesquisadores do Facebook quiseram descobrir como essa filtragem afetava as opiniões políticas que víamos. Se o modelo “filtro” se aplica para o compartilhamento de notícias políticas, então esperaríamos que o feed de notícias do Facebook fornecesse menos exposição das ideias contraditórias do que artigos compartilhados por amigos.

Os resultados deles eram muito claros. O efeito da filtragem era insignificante. Tanto liberais quanto conservadores ficavam expostos a somente um pouco menos de opiniões opostas por intermédio do filtro do que se o Facebook tivesse fornecido publicações aleatórias em seu feed. Nós *somos* mais propensos a ver artigos em nosso feed de amigos mais próximos, mas as opiniões políticas expressas *não* são mais extremas do que aquelas expressas por todos os nossos amigos. Grande parte do que vemos no Facebook não está de acordo com as nossas próprias opiniões. Além do mais, conservadores americanos, um grupo frequentemente acusado de ter a mente fechada, foram um pouco mais expostos a opiniões contrárias do que os liberais.

O teste da bolha-filtro foi publicado na prestigiosa revista científica *Science*. Trata-se apenas de um exemplo de como o Facebook, durante a primeira metade dessa década, levou a ciência a sério. Ele empregou pesquisadores de ponta para descobrir o efeito de seu produto — e esses pesquisadores pensaram grande.

Na campanha da eleição para o Congresso de 2010, uma mensagem foi mostrada a 60 milhões de usuários do Facebook nos Estados Unidos, colocada no topo dos feeds de notícias, ajudando-os a encontrar os lugares de votação deles e fornecendo um botão para indicar se já haviam votado. A mensagem continha fotos de amigos de usuários que já haviam clicado no botão “Eu votei”.

James Fowler, pesquisador da Universidade da Califórnia, decidiu medir o efeito dessa mensagem. Trabalhando juntamente com o Facebook, ele e seus colegas criaram um grupo menor de usuários, de cerca de 600.000

pessoas, para quem mostraram a mesma mensagem sem os rostos dos amigos que haviam votado.⁷ A hipótese deles: seria a natureza social das mensagens que criaria um efeito nos usuários. Ao olhar os rostos de nossos amigos que já haviam votado e nos dar a chance de lhes dizer que estamos com eles, somos encorajados a sair e votar.

A hipótese estava correta. Os usuários que viram a mensagem não social foram ligeiramente menos propensos a votar do que os usuários que viram quais de seus amigos haviam votado. A diferença da probabilidade de voto era de somente 0,34%. Mas mesmo uma pequena mudança no comportamento de voto pode fazer uma grande diferença no número de votantes. Ao comparar os usuários do Facebook com os registros eleitorais, James e seus colegas estimaram que eles tinham criado pelo menos 60.000 novos eleitores como resultado direto de terem visto a mensagem, e pelo menos um adicional de 280.000 que haviam votado como resultado do contágio social criado pela mensagem. Uma pequena cutucada fez uma grande diferença no número de pessoas participando da democracia.

O estudo político mostrou que as mensagens do Facebook poderiam ter um efeito no nosso cotidiano. O próximo passo, fornecido pelo cientista de dados do Facebook Adam Kramer, juntamente com dois pesquisadores de Cornell, foi analisar o efeito emocional global das mensagens. Eles conduziram um experimento completo. Os pesquisadores removeram entre 10% e 90% de publicações positivas do feed de notícias de cerca de 115.000 usuários, mostrando-lhes mais publicações negativas do que o costume. Publicações felizes de famílias se divertindo ou animais de estimação engraçados foram temporariamente removidas do feed de notícias dos usuários. Os pesquisadores mediram, então, o comportamento de publicação desses usuários em comparação com o grupo controlado que recebeu seu feed de notícias usual. A questão científica era como a remoção de publicações positivas afetaria as publicações que os usuários, que haviam sido submetidos a isso, fariam.

Na época em que esse artigo foi publicado, surgiram preocupações por parte dos membros da comunidade científica e um enorme protesto foi feito

na mídia sobre a ética envolvida no experimento. A objeção era que o Facebook estaria manipulando os feeds dos usuários e os “enganando” sem o consentimento deles. Isso poderia ser considerado, sob certas diretrizes regulamentais, antiético. Mas o debate on-line e na mídia estava menos preocupado com os detalhes de qual regra ética se aplicava ao experimento em particular, e muito mais preocupado com o fato de o Facebook estar nos manipulando.

Considere este protesto um tanto confuso. Claro que o Facebook está manipulando as informações às quais os usuários têm acesso! Esta é a ideia central desse modelo de negócios. Manipulação do feed de notícias é o algoritmo do Facebook que Adam Mosseri se orgulha de apresentar a líderes empresariais. A empresa está sempre ajustando o que lhe é mostrado de modo a fazê-lo utilizar o site com mais frequência. O Facebook é completamente aberto em relação a isso e até mesmo lhe permite ajustar sua vivência.

O resultado surpreendente do experimento de Adam Kramer foi que a manipulação do feed de notícias resultou em *quase nenhum efeito* nas emoções dos usuários. O estudo mostrou que o percentual de palavras positivas usadas por pessoas para quem as notícias positivas haviam sido removidas de seus feeds foi um pouco abaixo dos 5,15%, enquanto, no grupo de controle, com o feed padrão, ficou próximo dos 5,25%. Essa é uma diferença de 0,1%. Para termos uma ideia de quão pequeno um efeito de 0,1% pode ser em termos de publicações no Facebook, suponha que você seja um usuário relativamente ativo que publica cerca de 100 palavras por dia. Uma diminuição de 0,1% significa que, durante os próximos 10 dias, você irá escrever um total global correspondente a uma palavra positiva a menos. Talvez na próxima quarta-feira você descreva um filme a que tenha assistido usando “OK” em vez de “bom”.

Os resultados foram ainda mais fracos no caso das palavras negativas. O tamanho do efeito foi de 0,04%, correspondendo a aproximadamente uma palavra negativa a mais por mês. É difícil imaginar que qualquer outra pessoa iria notar o “chata” adicional que você adicionou à descrição de uma

reunião de trabalho, ou o fato de que você ficou “desapontado” com seu time de futebol por ter perdido um jogo importante.

Enquanto os jornais estavam debatendo os perigos de o Facebook estar realizando experimentos em nós, eles ignoraram o fato de os resultados serem desprezíveis. Isso foi, em parte, culpa de Adam Kramer e seus colegas. O título do artigo deles era: “Evidência experimental de contágio emocional em larga escala por intermédio das redes sociais.” O quê?! Em retrospectiva, esse título, que sugere que as emoções se espalham como o vírus Ebola por meio das nossas redes sociais, não era uma descrição particularmente acurada. O que eles haviam de fato encontrado teria sido mais bem descrito como um aumento mínimo, ainda que estatisticamente significativo, na expressão de emoções positivas como um resultado de manipulação em larga escala dos feeds de notícias dos usuários do Facebook.

Há muito exagero com o Facebook e seu efeito em nossa vida. Mas o que me deixou mais perplexo, após uma reanálise cuidadosa dos grandes estudos feitos e após conversar com os pesquisadores envolvidos, foi que os resultados eram quase sempre distorcidos ou exagerados quando transmitidos pela mídia.

O exagero dissonava com meu próprio entendimento da ciência. Sim, o Facebook poderia inflar uma pequena bolha no dia de uma eleição que fizesse um número um pouco maior de pessoas votar. Sim, poderia esvaziar ligeiramente nossa bolha emocional nos mostrando apenas publicações depressivas. E, sim, o Facebook não provê notícias que sejam inteiramente representativas da ampla variedade de opiniões expressadas pelo mundo inteiro. Mas nenhum desses seria um efeito transformador de vida. O efeito do Facebook em nossa vida é muito pequeno comparado ao efeito de nossas interações diárias com pessoas na vida real.

Consegui perceber que a analogia da bolha foi útil. Michela e seus colegas na Itália mostraram o quão insulares certos grupos poderiam se tornar. Mas a fragilidade da teoria da bolha, quando aplicada a grupos maiores de pessoas com interesses mais amplos, estava começando a me incomodar. A mídia social estava repleta de bolhas criadas por algoritmos, e

havia algumas teorias da conspiração bem malucas flutuando por intermédio delas. Mas, para a maioria de nós, parecia haver uma fuga das bolhas. O que foi que evitou que ficássemos presos? Se o algoritmo “filtro” do Facebook é capaz de, em teoria, nos aprisionar em um ponto de vista, como escapamos na realidade? Por que as nossas emoções ficam largamente inalteradas durante os longos períodos de tempo que passamos utilizando a mídia social?

A única forma que encontrei para responder a essas perguntas foi imergir na minha própria bolha on-line e ver se conseguia emergir dela. Não foi algo que eu estava particularmente entusiasmado em fazer. Tenho uma bolha que prezo muito e não gostaria de estourá-la. Mas não tinha como inventar mais desculpas. Eu tinha que descobrir do que minha própria bolha era feita.

O FUTEBOL IMPORTA

Meu uso das redes sociais gira em torno do futebol. Tenho uma conta no Twitter, @Soccermatics, na qual converso com colegas nerds matemáticos do futebol sobre o jogo, e a matemática e as estatísticas usadas para analisá-lo. Essa pequena parte do universo Twitter é tanto uma bolha-filtro como uma câmara de eco. Eu sei que o Twitter filtra meu feed de modo que os maiores geeks estão no topo da minha página principal. Outras contas sobre a estatística do futebol, tais como @deepxg, @MC_of_A e @BassTunedToRed, falam coisas que me interessam e o Twitter se certifica de que eu veja seus tweets primeiro. Minha tendência em seguir outros geeks que pensam como eu cria uma bolha na qual todos concordamos sobre a necessidade de a matemática e o futebol trabalharem juntos.

Ocasionalmente, usuários menos nerds do Twitter (normalmente fãs do Chelsea) que não apreciaram minha mais recente visualização de rede de passes aparecem na minha câmara de eco e me dizem que eu deveria, por exemplo: “me masturbar com minha planilha”, ou “em vez disso, vá conhecer umas garotas, seu virgem”. Mas, como essas declarações não contêm informação científica, normalmente as ignoro ou, educadamente, conecto o fã do Chelsea a um artigo que explica os conceitos básicos que estão por trás das análises do futebol.

Admito livremente viver em uma bolha analítica do futebol. Mas percebi algo interessante sobre essa bolha. Alguns de nós, assim como eu, somos fãs do Liverpool. Alguns torcem para o Everton. Alguns torcem para o Manchester United e outros, para o City. Alguns até torcem para o Chelsea. Há fãs do Real Madrid, fãs do Barcelona, fãs do futebol italiano e do futebol alemão. E há analistas que tuítam sobre futebol africano, futebol asiático e dos Estados Unidos. Em vez de habitar uma bolha para cada time ou para cada liga, nós, geeks matemáticos do futebol, vemos e compartilhamos informações de todas as formas de futebol. No momento, sei tanto sobre o Manchester City quanto sobre o Liverpool. Aprendi sobre a Liga Principal de Futebol dos Estados Unidos (e Canadá, como descobri recentemente) e até mais sobre a liga feminina, NWSL. Já dei, também, uma fugida da bolha do futebol para dar uma espiada na do hóquei no gelo, na do futebol americano e do basquete.

A política também se infiltra na minha bolha. Há, muitas vezes, um forte ethos socialista da classe trabalhadora entre os fãs do Liverpool que eu sigo. Mas também sigo jornalistas de impressos de direita do Reino Unido e outros meios de comunicação. Muitos dos analistas dos Estados Unidos que eu sigo tendem a apoiar os democratas e retuitar artigos contra Trump, mas também sigo técnicos trabalhando nas categorias de base por todo o país, simpatizantes de ideais republicanos e motivados pela fé cristã. Um analista em particular, @SaturdayOnCouch, me mantém entretido tanto com seus mapas de calor do futebol alemão quanto com sua trollagem meio amigável de analistas de futebol liberais e anti-Trump.

Vejo pedidos para o Manchester United montar um time de futebol feminino, campanhas dos fãs do Leicester City para acabar com abusos homofóbicos nas arquibancadas e apelos dos fãs do Borussia Dortmund para a Alemanha receber refugiados da Síria. Vejo campanhas políticas e histórias sobre como a insatisfação entre as classes trabalhadoras americanas levou à ascensão de Donald Trump.

Imaginar a política se infiltrando em nossa bolha de mídia social é uma maneira útil de visualizar como nos deparamos com diferentes pontos de

vista políticos. Bob Huckfeldt, pesquisador da Universidade da Califórnia, em Davis, passou mais de trinta anos investigando nossas discussões políticas. Durante a campanha Bush-Clinton de 1992, Bob e seus colegas pediram aos simpatizantes dos dois partidos que listassem as afiliações políticas das pessoas com as quais haviam conversado sobre política. Eles relataram que 39% das discussões políticas foram com pessoas que não apoiavam o mesmo candidato à presidência que eles.¹ Essa estatística poderia ser parcialmente explicada pela tendência de os eleitores se lembrarem melhor das discussões com adversários, mas não inteiramente. Na corrida para a eleição de Bush-Clinton, pessoas de todos os Estados Unidos estavam regularmente discutindo sobre política com pessoas de ponto de vista oposto.

Bob enfatiza que “as pessoas levam sua vida dentro de vários contextos, definidos por múltiplas dimensões”.² Esportes e política são um exemplo de duas dimensões sociais diferentes. Durante o jogo, e em discussões on-line após o jogo, fãs do esporte conhecem pessoas com todo tipo de opiniões políticas diferentes. O mesmo acontece com música, livros, filmes, alimentação e fofocas sobre celebridades. Pessoas de vários contextos políticos diferentes se encontram on-line para falar sobre *A garota no trem* ou *Os vingadores* e acabam discutindo sobre política.

Será que a campanha de 1992 faz parte de uma época dourada de amizade entre colegas de trabalho e fãs do esporte conversando durante o almoço ou em bares sobre os prós e contras de dois candidatos? De jeito nenhum. Bob observou níveis parecidos de discussões que conectam a divisão política na Alemanha e no Japão. Ele encontrou o mesmo resultado no que era considerada, até recentemente, a campanha presidencial mais controversa dos últimos cinquenta anos: George W. Bush *versus* Al Gore em 2000. Mais de um terço das discussões políticas entre os eleitores pesquisados durante essa eleição aconteceu entre simpatizantes de partidos opostos.

Essas discussões políticas não são apenas: “Uhu!, eu sei mais do que você.” Bob e seus colegas também testaram o conhecimento político das

pessoas em seus estudos e descobriram que era aos especialistas que os outros recorriam quando queriam saber mais sobre um determinado assunto. Independentemente de seus pontos de vista, aquelas pessoas que sabiam mais sobre política eram as mais propensas a serem questionadas sobre suas opiniões.

A dúvida persiste, contudo, sobre se a mídia social transformou ou não — nos 18 anos desde Bush *versus* Gore — a forma por meio da qual trocamos informações sobre assuntos políticos. A pesquisa de Bob foi feita numa era anterior à chegada do Twitter, Facebook e Reddit. Recentemente, tem-se especulado bastante que as coisas agora estão realmente diferentes.³ Devido às preocupações sobre polarização on-line durante campanhas como a votação do Brexit e a eleição presidencial dos Estados Unidos, eu quis verificar se o trabalho de Bob ainda se aplicava nos dias de hoje ou se as câmaras de eco haviam assumido o controle.

Decidi testar essa ideia, começando comigo mesmo.

A Figura 12.1 mostra parte da minha rede social no Twitter. Cada círculo da figura é uma pessoa que sigo e que me segue de volta. Eu sou o círculo central, com todas as linhas conectando-se a mim (uma vez que sou o ponto focal de minha própria rede). Outras linhas desenhadas entre pares de pessoas indicam que eles também satisfazem a relação seguir/seguir de volta. Essa figura só inclui usuários “típicos” do Twitter, que eu defino como pessoas que seguem menos que 1.000 pessoas e são seguidas por menos de 1.000 pessoas. Acadêmicos de muito prestígio, celebridades e personalidades da mídia estão, portanto, excluídos.

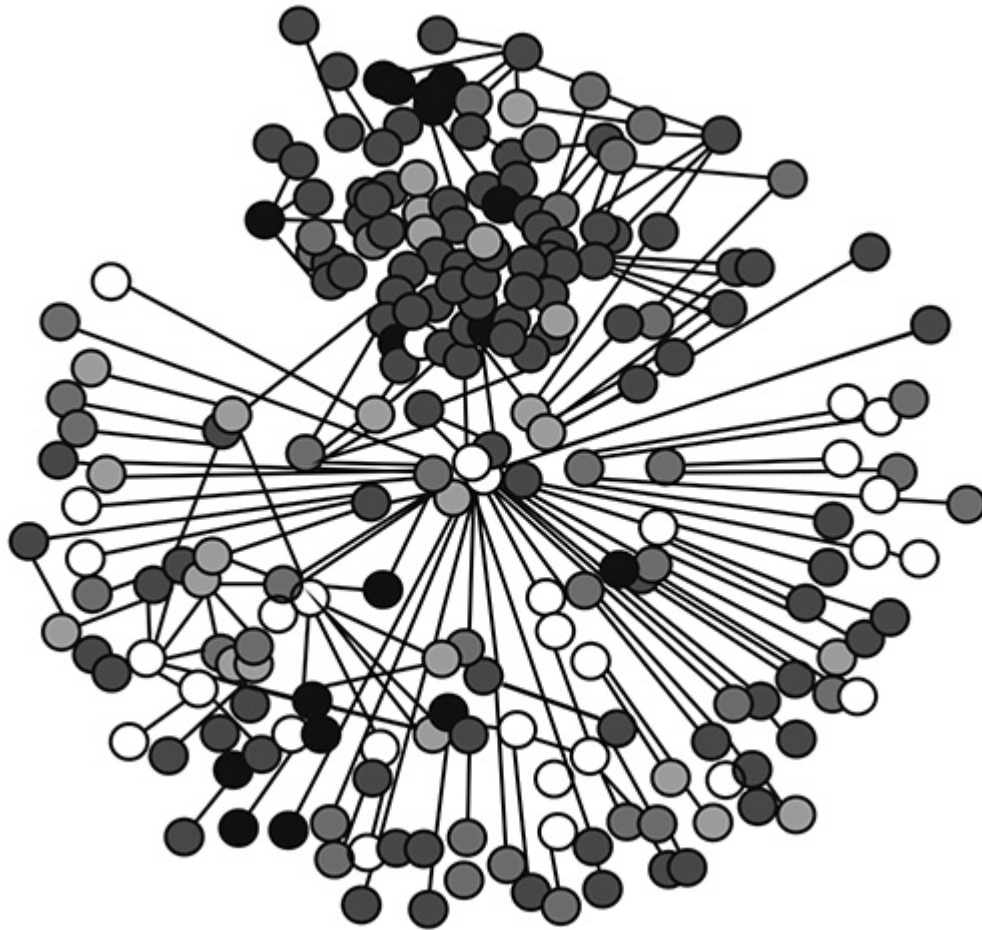


Figura 12.1 Minha rede social do Twitter. Todos os usuários do Twitter (com menos de 1.000 seguidores) que me seguem e que eu sigo de volta estão representados com círculos. Eu sou o círculo central. As linhas conectam pessoas que seguem umas às outras. Pessoas cujos círculos são pretos têm um baixo grau de separação dos três jornais britânicos que publicaram mais artigos com uma inclinação “pró-permanência” na campanha de votação do Brexit (Daily Mirror, Guardian e Financial Times). Círculos brancos indicam um baixo grau de separação dos jornais com uma inclinação maior “pró-saída” (The Times, Daily Star e Telegraph, The Sun, Daily Mail e Daily Express).⁴ Os círculos cinza indicam um “grau de separação” entre esses dois extremos. Figura criada por Joakim Johansson.

Existe uma estrutura muito nítida na minha rede. No topo da Figura 12.1 há um aglomerado de pessoas que estão conectadas entre si bem como a mim. Essas pessoas são cientistas. Elas se aglomeram para compartilhar as mais recentes descobertas, rir sobre seus dias de doutorado e reclamar sobre a situação do fomento científico. Esse grupo é o melhor tipo de câmara de

eco: nos reafirmamos uns aos outros, apoiamos uns aos outros e compartilhamos as fofocas mais recentes.

As outras conexões de minha rede estão mais espalhadas. Trata-se de pessoas que não se conhecem entre si, mas que me conhecem. É aqui que muitos dos meus colegas fãs de futebol podem ser encontrados. Minha escolha quanto a quem seguir sobre futebol é mais aleatória do que na ciência. Algumas vezes compartilho tweets para cá e para lá sobre um jogo ou algum jogador, começo a gostar da conversa e, então, decido seguir esse usuário com quem estava interagindo. Provavelmente essa pessoa não conhece as outras pessoas aleatórias que eu segui desse modo.

A fim de perceber as influências políticas agindo nos meus amigos do Twitter, um de meus alunos de mestrado, Joakim Johansson, teve uma ideia muito legal. Ele usou a questão do Brexit como o assunto político determinante na minha rede, uma vez que a maior parte de meus seguidores encontra-se no Reino Unido e este é um assunto quente no meu feed. Joakim sombreou as pessoas que eu conhecia em seus “graus de separação” dos jornais do Reino Unido. As pessoas mais próximas, em termos de seguidas no Twitter, aos jornais que eram simpatizantes à campanha de permanência foram representados com círculos mais escuros, enquanto aqueles apoiando a campanha de saída foram representados por um círculo mais claro.

Para medir o grau de separação, Joakim olhou para quantos usuários do Twitter precisariam ser atravessados até se chegar a cada um dos nove jornais mais proeminentes do Reino Unido. Se uma pessoa seguir o *Guardian*, então seu grau de separação do jornal é zero. Se ela não seguir o *Guardian*, mas seu amigo seguir, então o grau de separação é um. Se o amigo de um amigo seguir o jornal, então o grau de separação é dois, e assim por diante. Essa é a medida padrão de grau de separação normalmente associada ao número seis. Dizem que há no máximo (apenas ligeiramente incorreto) seis graus de separação entre nós e qualquer outra pessoa no planeta.⁵

As pessoas mostradas em minha rede social que aparecem como um círculo mais escuro possuem um baixo grau de separação de jornais “pró-permanência” na corrida à votação, enquanto um círculo mais claro indica um baixo grau de separação dos jornais “pro-saída” (Figura 12.1). Aqui vemos uma ligeira diferença entre meu aglomerado de amigos cientistas — mais próximos de jornais como o *Guardian*, que se opunha ao Brexit — e meus amigos fora da ciência, que possuem uma variedade mais ampla em seu sombreamento. O mais impressionante, e que me surpreendeu mesmo após ter estudado os trabalhos de Bob Huckfeldt, foi a variação nos graus de separação entre meus amigos e jornais de saída/permanência. O futebol e o Twitter em geral me expõem a uma grande variedade de opiniões.

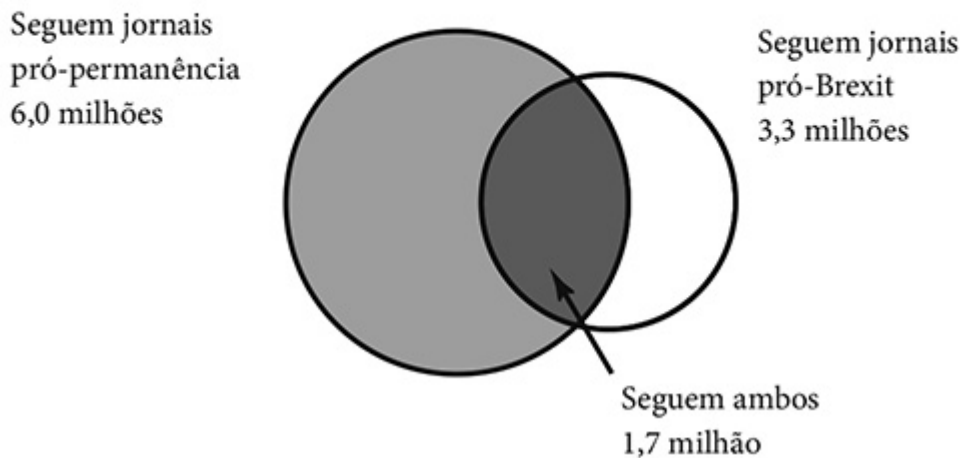
Eu mal posso ser considerado representativo, de modo que, em sua tese, Joakim criou redes para centenas de usuários diferentes, cada um deles seguia pelo menos um jornal do Reino Unido.⁶ As redes sociais desses usuários consistiam, tipicamente, em um a dois aglomerados coesos de amigos que seguiam principalmente uns aos outros. Algumas vezes os usuários desses aglomerados estariam todos próximos a jornais pró-permanência, outras vezes, pró-saída. Entretanto, além desses agrupamentos mais partidários, havia sempre ramos isolados de amigos que continham um conjunto completamente diferente de conexões. Esses ramos asseguravam as conexões entre usuários com opiniões políticas diferentes.

Para os usuários do Twitter que seguem só jornais pró-permanência, apenas 13% de seus amigos seguem jornais exclusivamente pró-saída. Esse é um número muito pequeno, comparado aos 54% que seguem exclusivamente os jornais pró-permanência. Entretanto, outros 33% de seus amigos seguem jornais pró-saída e pró-permanência. Assim, em termos de graus de separação, a maioria de nós está apenas a alguns poucos passos de um jornal que se opõe às nossas opiniões. Contas do Twitter que são exclusivamente pró-saída possuem, em média, apenas 1,2 grau de separação do *Guardian*. Contas que são exclusivamente pró-permanência estão separadas apenas por uma média de 1,5 conta do *The Sun*.

Uma exceção a esta regra é o tabloide sensacionalista *Daily Star*. Ele está, em média, a 2,2 passos de contas pró-permanência, mas também está, em média, a 2,2 passos de contas pró-saída. O *Daily Star* é conhecido por publicar o tipo de teorias da conspiração dúbias que se encaixariam em um dos estudos de Michela Del Vicario. O restante de nós permanece, de alguma forma, desconectado dessas opiniões.

Usuários individuais do Twitter frequentemente querem ouvir ambos os lados da história. Para ilustrar isso, fiz um diagrama de Venn dos jornais seguidos no Twitter logo após a eleição no Reino Unido para sair da União Europeia. A Figura 12.2a mostra que há uma grande sobreposição entre seguidores do Brexit e jornais pró-permanência. É comum, para as pessoas, seguir tanto o *Guardian* quanto o *Telegraph*, por exemplo. Essa sobreposição fica reduzida se olharmos para tabloides pró-saída e grandes jornais pró-permanência (Figura 12.2b).

(a)



(b)

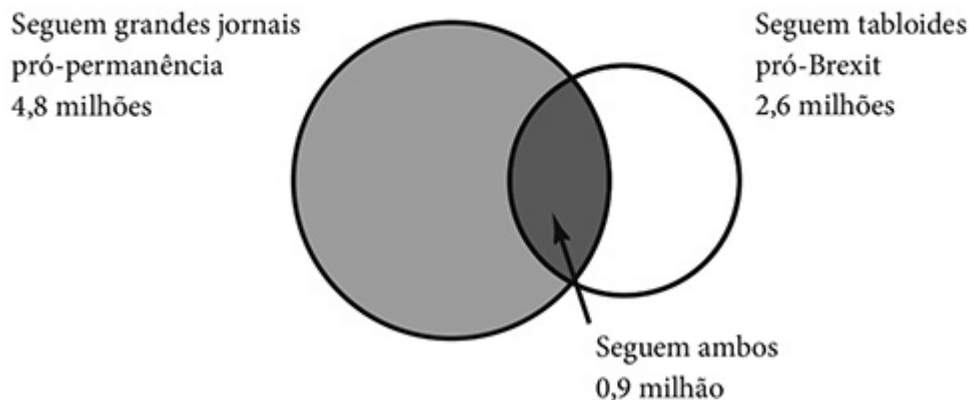


Figura 12.2 Diagrama de Venn de como as pessoas seguem jornais do Reino Unido no Twitter. Dividi os jornais em grandes jornais “pró-permanência” (Guardian, Independent e Financial Times), grandes jornais pró-Brexit (The Times e Telegraph), tabloides pró-permanência (Daily Mirror) e tabloides pró-saída (Sun e Mail Online). (a) Número de pessoas que seguiam jornais pró-permanência, jornais pró-Brexit e ambos; (b) Número de pessoas que seguiam grandes jornais pró-permanência, tabloides pró-saída e ambos.

A divisão em seguir jornais no Reino Unido é menos relacionada às opiniões políticas e mais à antiga divisão entre tabloides e grandes jornais. Muitos leitores do *Guardian* já seguem o *Telegraph*. Assim, se você ler o *Guardian* e realmente quiser ampliar seus horizontes, siga, então, o *Daily Mail*. Vá lá. Tente.

Nas discussões políticas on-line, tendemos a passar mais tempo discutindo coisas que não gostamos do que coisas que gostamos. Um estudo das primárias presidenciais descobriu que Hillary Clinton era mencionada mais vezes no Twitter pelos simpatizantes dos republicanos do que dos democratas.⁷ Da mesma forma, os democratas tuitavam mais sobre Donald Trump do que os republicanos. Simplesmente não conseguimos resistir à tentação de dar enfoque a críticas ao líder do outro lado.

O algoritmo de classificação do Twitter muda esses sentimentos ocultos. A cientista de computação Juhi Kulshrestha e suas colegas descobriram que buscas por Hillary Clinton no Twitter, durante a eleição, revelavam tweets

que eram mais simpatizantes a ela do que refletia o sentimento geral dos tweets feitos. Buscas por Donald Trump, por outro lado, reforçavam a imagem negativa do candidato. Como no Reino Unido, os usuários do Twitter nos Estados Unidos têm uma leve tendência liberal, e a forma com a qual o Twitter filtra essa tendência serve para aumentá-la (ligeiramente).

Lendo a pesquisa no Twitter e conduzindo meus próprios experimentos cheguei a conclusões similares às aquelas alcançadas a partir dos trabalhos de Lada Adamic no Facebook. Para as pessoas que seguem jornais e se mantêm atualizadas sobre os problemas atuais, como eu, o Twitter e o Facebook não são particularmente câmaras de eco fortes. Esses sites de mídia social ajudam a difundir e compartilhar informações variadas, apesar de uma leve inclinação liberal. Ao todo, ouvimos muitas opiniões, algumas das quais gostamos, outras, não, mas todas elas nos mantêm informados sobre o mundo onde vivemos. Nossas conexões sociais de longo alcance nos mantêm fora de uma bolha.

Havia ainda uma preocupação grande que era repetidamente citada, tanto por pesquisadores com quem eu falava quanto em várias formas de opinião na mídia liberal. Essa preocupação não é sobre as pessoas que leem os jornais. É sobre as pessoas que estão desconectadas das notícias tradicionais, recebendo notícias de outras fontes potencialmente menos confiáveis. Os usuários do Twitter que Joakim e eu estudamos já eram seguidores de jornais do Reino Unido. Poderíamos discutir sobre a qualidade do jornalismo nessas diferentes publicações, mas já são meios de comunicação bem estabelecidos, governados por códigos de prática e pelas leis do Reino Unido.

Com Donald Trump usando o Twitter para distorcer a verdade, ao mesmo tempo acusando a mídia de fazer a mesma coisa, com sites como o Breitbart distorcendo histórias de formas enganosas e com páginas do Facebook espalhando rumores ultrajantes, havia uma preocupação legítima de que muitas pessoas não estavam recebendo as notícias de fontes confiáveis. Era muito fácil, para mim, me declarar, e a outros leitores de jornais, sem filtro, mas e quanto às pessoas que ignoram a mídia tradicional?

Eu já havia xeretado algumas bolhas de conspiração e o que encontrei foi bastante preocupante. Fake news: a proliferação de rumores não verdadeiros sobre líderes políticos era abundante e poderia estar influenciando outros grupos de pessoas que estavam fugindo da mídia tradicional.

De acordo com muitos dos artigos publicados pela mídia tradicional, é exatamente isso o que está acontecendo. Muitas pessoas estão supostamente desaparecendo em um mundo de fofocas de entretenimento, resultados de esportes, filmes curtos de animais de estimação e memes on-line. Um mundo onde as notícias reais são chatas e as fake news são uma fonte de entretenimento constante. Um mundo pós-verdade.

Estariam algumas pessoas vivendo num mundo pós-verdade? Eu precisava descobrir.

QUEM LÊ FAKE NEWS?

Master Bates e Seaman Staines.* Quando ouvi, ainda um estudante, no início dos anos 1990, o boato de que esses eram os nomes da tripulação do navio de um dos meus programas infantis preferidos, *Capitão Pugwash*, não questionei a informação. Pareceu de certa forma plausível que, uma vez que meus amigos e eu éramos tão inocentes na época em que o show foi transmitido nos anos 1970, ninguém tenha percebido a escolha dos nomes. O boato teve origem em um artigo do jornal *Guardian*, que nunca vi. Simplesmente aceitei que poderia ser verdade e passei adiante. Era engraçado.

Fake news. O criador de *Capitão Pugwash* processou o *Guardian* e ganhou.

Durante muitos anos depois do boato de Pugwash, eu ria com amigos sobre como havíamos nos deixado enganar. Parecia uma ideia excepcionalmente boba.

Não mais. Outra noite, meu filho e minha filha me contaram sobre algo chamado o “efeito Mandela”. Eu não fazia ideia do que se tratava.

Minha filha explicou fazendo uma pergunta: “Existe uma marca preta na cauda do Pikachu [um Pokémon]?”

“Sim”, eu disse. Achei que havia.

“Não tem”, Elise me disse. “Muitas pessoas acham que tem e incluem uma marca preta quando desenham o Pikachu. Mas não deveriam.”

O efeito Mandela é quando você acha que alguma coisa é verdadeira, mas não é.

Bem, sim. Entendi o conceito. Mas por que é chamado efeito Mandela? “Pesquise no Google”, disse minha filha.

Pesquisei. O Google até mesmo usou o autopreenchimento como a melhor sugestão quando digitei “Mandel... “. Havia, naturalmente, um youtuber à disposição para explicar toda a história. Ele explicou a cauda do Pikachu, a questão sobre o cara do Monopoly ter um monóculo ou não, o que o Darth Vader realmente disse a Luke Skywalker e outros exemplos de memórias falsas. Até aí, tudo bem.

O que me surpreendeu foi a história original sobre Mandela. De acordo com o youtuber, muitas pessoas acreditam que Nelson Mandela morrera na prisão, nos anos 1980. Ele e muitos outros como ele, incluindo meus próprios filhos, consideravam a morte de Nelson Mandela numa prisão o exemplo primordial de memória falsa. Mas não é. Não há qualquer evidência convincente de que um grande número de pessoas acredita que Mandela morrera numa prisão. Uma pesquisa rápida que fiz mostrou que toda a ideia poderia ser atribuída a uma única publicação num blog, escrita pela “consultora paranormal” Fiona Broome em 2010. Eu estava de volta ao mundo das teorias da conspiração invadindo a normalidade. Não havia evidências de o efeito Mandela ter jamais existido em sua forma original.

O efeito Mandela é por si próprio um efeito Mandela. Apesar do fato de milhares de pessoas acreditarem que Mandela morreu na prisão nunca ter acontecido, ele se tornou, no YouTube, o nome do fenômeno cientificamente estabelecido como memória falsa.

A indústria de fake news está em crescimento. Os vídeos do YouTube sobre o efeito Mandela tiveram milhões de visualizações. Os youtubers são pagos pelos anúncios a que temos de assistir antes de os vídeos começarem. Eles acumulam seus “gostei” e passam para a próxima teoria conspiratória. Esses vídeos em particular eram apenas um pouco enganosos,

especificamente no assunto Mandela, mas outros sites de notícias falsas têm conteúdo muito mais dúbio.

Durante a eleição presidencial americana e no ano seguinte, as fake news realmente decolaram. O editor fundador do BuzzFeed, Craig Silverman, fez uma lista das grandes histórias sobre a corrida Trump *versus* Clinton.¹ A maior parte delas era pró-Trump, com manchetes do tipo: “O Papa endossa Donald Trump”; “Os e-mails de Hillary sobre ISIS acabaram de vazar”; “Agente do FBI que investigava Hillary foi encontrado morto”. Mas também havia manchetes anti-Trump: “RuPaul disse que Donald Trump o apalpou.” Algumas dessas histórias vieram de sites que foram criados para serem satíricos, outros eram administrados por simpatizantes de direita. Um número grande de histórias teve origem em uma cidade pequena da Macedônia, onde um grupo de jovens foi pago pelos anúncios mostrados nos sites. Sem sequer considerar se as histórias eram verdadeiras ou não, eles as colocaram no Facebook, uma depois da outra, na esperança de que pelo menos alguma história viralizasse e fizesse com que eles ganhassem dinheiro.

Donald Trump rotulou o *New York Times*, a CNN e outras fontes jornalísticas tradicionais de fake news, por causa do que ele entende como sendo uma cobertura unilateral de seu mandato presidencial. Trump pode usar essa definição se ele quiser. Minha definição é mais rígida: fake news são notícias demonstravelmente falsas, não apenas posicionamentos políticos. Notícias falsas consistem em histórias escolhidas por sites que investigam farsas como Snopes e PolitiFact, e mostradas como factualmente incorretas. Baseado nessa definição, havia pelo menos 65 sites de fake news durante a eleição. O site de extrema direita Breitbart se encontra delicadamente equilibrado na borda da minha definição.

A questão não é se fake news existem ou não. Há poucas dúvidas a respeito. A questão é o quanto elas influenciam nossos pontos de vista políticos. Será que vivemos num mundo pós-verdade?

O único modo de testar apropriadamente a hipótese da pós-verdade é conduzir uma pesquisa e analisar os dados. Foi exatamente isso que os

economistas Hunt Allcott e Matthew Gentzkow fizeram.² Eles quiseram medir o efeito de fake news na corrida para a eleição presidencial americana de 2016. Eles apresentaram aos participantes de uma pesquisa on-line uma série de fake news, incluindo:

“Funcionários da Fundação Clinton foram condenados por desviarem verbas para comprar bebidas alcoólicas para festas caras no Caribe”

“Mike Pence disse que ‘Michelle Obama é a primeira-dama mais vulgar que já tivemos’”

“Documentos vazados revelam que a campanha de Trump planejou um esquema para dar carona para eleitores democratas até as urnas, mas então os levava para o lugar errado”

“Num comício poucos dias antes da eleição, o presidente Obama gritou com um manifestante que apoiava Donald Trump”

As reações dos participantes a essas notícias foram comparadas às reações que eles tinham quando lhes eram mostradas notícias verdadeiras.

Duas perguntas foram feitas aos participantes: se eles tinham ouvido a notícia relatada e se acreditavam nela. Na média, cerca de 15% das pessoas lembraram de ter ouvido essas notícias falsas, comparados aos 70% que tinham ouvido as notícias verdadeiras. Poderíamos concluir que 15% de chance de se ouvir qualquer notícia falsa é bastante grande. Se quiser, faça você mesmo o teste agora. Quantas das notícias falsas listadas anteriormente você lembra de ter ouvido durante a eleição?

Se você se lembrou de mais da metade, então acho que temos um problema. Somente dois desses boatos foram de fato espalhados on-line. Os outros dois foram mostrados por Hunt e Matthew aos participantes como uma forma de placebo. Eles mesmos inventaram a primeira e a terceira notícia. No experimento, 14% das pessoas disseram que tinham ouvido essas notícias falsas. Isso não foi significativamente diferente dos 15% que

disseram que haviam ouvido as notícias de fato falsas. Mesmo pouco tempo depois da eleição, os participantes não conseguiram lembrar direito de quais coisas falsas eles haviam visto on-line.

Combinando esses resultados com uma análise de como as notícias se espalham no Facebook e o impacto relativo de sites de notícias falsas comparado aos sites de notícias tradicionais, Hunt e Matthew fizeram um cálculo “num guardanapo” para mostrar que, na melhor das hipóteses, o americano médio seria capaz de lembrar de uma ou duas notícias falsas quando ele fosse às urnas, e que era improvável que acreditasse nelas.

Quando o contatei, Matthew estava relutante em fazer uma conclusão definitiva de que as fake news não tiveram influência na eleição. Ele me disse: “Não podemos estimar como ler uma notícia/anúncio afeta de fato a maneira de as pessoas votarem.” Conectar exposição às notícias com voto requer mais estudos.

Mesmo se formos tão cautelosos como Matthew sugere, eu simplesmente não conseguia enxergar como o efeito de fake news fazia sentido. Apesar da vitória apertada de Donald Trump na eleição, o cálculo “feito num guardanapo” de Hunt e Matthew implica que fake news não passam de um barulho sem significado.

A ineficácia de fake news proporciona um cenário bastante diferente daquele insinuado pelo conceito de um mundo pós-verdade. Sim, há muitas fake news escritas, mas elas são esquecidas de imediato e raramente são levadas a sério.

Bob Huckfeldt me disse que uma das principais mensagens que ficam de sua própria pesquisa é de que “cidadãos interessados e informados sobre política têm sempre sido particularmente influentes [sobre seus pares]”. Existe pouco motivo para se acreditar que isso mudou, só porque um bando de adolescentes da Macedônia tem inventado histórias a fim de faturar com anúncios.

Um dos exemplos mais proeminentes de fake news recentes surgiu no dia seguinte à vitória presidencial de Donald Trump. Os americanos que digitaram “contagem final da eleição” no Google Notícias tiveram uma

grande surpresa. O primeiro resultado da busca era um artigo de um site, 70 News, que declarava: “Números finais da eleição 2016: Trump ganhou tanto no voto popular quanto no colégio eleitoral.”³ A declaração estava incorreta; apesar de Hilary Clinton ter perdido a eleição, ela havia ganho nos votos populares por vários milhões de votos.

Quando entrei no site 70 News encontrei uma série de opiniões que não havia visto antes. O artigo da “contagem final da eleição” afirmava que mais de três milhões de imigrantes ilegais haviam votado na eleição. Supondo que a maioria desses “ilegais” votaram em Hillary Clinton, 70 News concluiu que Trump havia, de fato, ganho no voto popular.

A fonte dessas afirmações sobre as irregularidades na votação parecia ser de um tweet e era óbvio que o 70 News não tinha credibilidade alguma. Mas não consegui evitar o que acabou se tornando uma meia hora um tanto divertida estudando os outros “fatos” do site. O 70 News percebeu que Donald Trump faria 70 anos, 7 meses e 7 dias no dia 21 de janeiro de 2017, o dia seguinte de sua posse e seu primeiro dia inteiro no poder. Vinte e um é igual a $7+7+7$. São muitos setes e, na interpretação do 70 News, 777 é o número de Deus. Então Trump estava profetizado pela data.

A profecia Trump 777 acabou sendo verdadeira, matematicamente falando, pelo menos. A profecia era um pouco menos provável de valer, e nada de sobrenatural aconteceu na posse de Trump que a sustentasse. Infelizmente, essa era a única parte curiosa sobre numerologia no site, que, fora isso, consistia em várias declarações racistas, teorias da conspiração e propaganda de direita, a maior parte proveniente de mídias sociais.

Como uma página do 70 News recebe uma alta classificação no Google? Para descobrir, usei primeiro o serviço SharedCount para ver quais sites estavam compartilhando a publicação do 70 News. O site informou um incrível meio milhão de compartilhamentos no Facebook, enquanto outros sites, como Twitter, Google+ e LinkedIn, tinham apenas poucas centenas de compartilhamentos cada.⁴ O Facebook era claramente o culpado, o que foi confirmado quando procurei pelo link no site. Ele havia sido compartilhado várias vezes, por vários indivíduos diferentes. A origem desse sucesso pode

ser rastreada até um pequeno grupo de sites americanos de direita. As páginas do Facebook “Os veteranos americanos são amados”, “Rede de fãs de Trump” e “Will B. Candid” tinham todas compartilhado o link, frequentemente com um comentário, e foram todas, por sua vez, compartilhadas por muitos seguidores.

Essa disseminação da publicação do 70 News seguiu o mesmo caminho que a das teorias da conspiração italianas estudadas por Michela Del Vicario. Começou dentro de uma câmara de eco de simpatizantes de extrema direita, mas se tornou tão grande que ultrapassou os limites do grupo usual de opiniões extremistas. Vários simpatizantes de Trump acharam que ela “fazia sentido” e a compartilharam. Outros simpatizantes questionaram a validade da publicação, enfatizando que, de qualquer jeito, não importava, já que o lado deles havia ganhado. Como nas teorias da conspiração, fake news têm seus perpetradores e seguidores, mas uma vez que elas crescem, são frequentemente desafiadas. O algoritmo do Google deixou-se levar com a celebração da vitória e colocou a publicação como notícia de primeira página.

Pensei em formigas num balde. Isso pode não ser o que vem à cabeça de todo mundo que acabou de ler o 70 News, mas penso muito sobre as formigas. Elas são animais fascinantes, com todo tipo de mecanismos incríveis de comunicação. Elas deixam rastros químicos, conhecidos por feromônios, para mostrar o caminho até a comida, para marcar seu território e até mesmo dizer às suas parceiras de ninho quais lugares devem evitar. Uma colônia de formigas é um superorganismo. Formigas podem construir ninhos com mais de um metro de altura e enterrados a vários metros abaixo da terra, elas podem construir uma rede de cadeias de suprimento cobrindo vários quilômetros quadrados e algumas espécies até mesmo plantam e cultivam sua própria comida.

Mas se você colocar formigas num balde, elas se tornam bem estúpidas. Ouvi sobre o experimento do balde pela primeira vez por intermédio do biólogo Nigel Franks da Universidade de Bristol. Ele e seus colegas estavam no Panamá em 1989 quando tiveram a ideia, inspirada por estudos clássicos

da década de 1940. Eles colocaram formigas guerreiras em um balde, cujas bordas estavam cobertas por um material que as impedia de escapar, e as filmaram de cima. As formigas ficaram andando em círculos e círculos, e, ao fazer isso, andavam cada vez mais rápido. À medida que andavam, elas depositavam o feromônio químico, o que fazia com que as formigas atrás delas achassem que havia algo interessante para se encontrar mais à frente. Isso fez com que todas as formigas aumentassem a velocidade. Um circuito de feedback social foi criado e não demorou muito até que elas estivessem todas em velocidade máxima.

Estudei o fenômeno de feedback social levando à estupidez numa vasta gama de espécies. Meu colaborador, Ashley Ward, da Universidade de Sydney, mostrou que era possível enganar peixes esgana-gata para nadarem na frente de um predador se eles achassem que outro peixe tivesse feito isso primeiro. Em um experimento sobre navegação de pombos em bando, de Dora Biro, da Universidade de Oxford, estudamos um pássaro que Dora mais tarde apelidou de “pássaro doido” porque ele liderou outros pássaros em longos caminhos tortuosos até em casa, mesmo quando o pássaro que o estava seguindo conhecia um caminho mais curto. Todos esses são exemplos de feedback social em que regras de interação fazem com que grupos se confundam e façam coisas tolas.

A questão é o que nós, cientistas, concluímos quando vemos um feedback social levando à estupidez coletiva. Podemos ficar tentados a concluir que copiar os outros ou seguir os outros é prejudicial à sobrevivência animal. Podemos ficar tentados a desenvolver uma teoria chamada “bolha-feedback”, que descreve todos os exemplos obscuros de animais fazendo coisas estúpidas em grupo. Podemos descrevê-los como vivendo num mundo “pós-sobrevivência”, onde não se importam com o fato de que vão morrer ao ficar andando em círculos.

Tal teoria ainda não se materializou por um bom motivo: também vemos animais em grupo fazerem juntos várias coisas inteligentes. Nigel Franks fez seu experimento do balde a fim de entender melhor como formigas guerreiras conduziam grandes invasões em enxame que expurgavam o solo

da floresta de toda comida possível.⁵ No mesmo conjunto de experimentos em que Ashley Ward enganou peixes mostrando a eles uma réplica de peixe nadando na frente de um predador, ele também descobriu que (quando não tentamos enganá-los) cardumes são muito melhores do que indivíduos em identificar e evitar predadores.⁶ Dora Biro mostrou que seus pombos aprendem caminhos juntos e que, normalmente, acham o caminho de casa mais rápido em grupos.⁷ Colônias de formigas, bandos de pássaros e cardumes são inteligentes.

O caso do 70 News é um exemplo do Google e do Facebook sendo pegos em circuito de feedback. Eles experimentaram um aumento nas buscas e compartilhamentos de determinado termo e seus algoritmos deram proeminência a ele. Mas erros como esse são raros porque essas empresas incentivam sua erradicação. Ambas as empresas tentam evitar alimentar circuitos que promovem opiniões negativas sobre tópicos dominantes, exatamente por causa da publicidade negativa que recebem. Comecei uma conversa no Messenger do Facebook com um ativista de extrema direita que trabalha na página do Facebook de Will B. Candid, que havia sido essencial no compartilhamento da publicação do 70 News. Ele não foi particularmente cândido quando lhe perguntei sobre o que eles fazem, mas fez questão de me dizer uma coisa: “Os algoritmos que o Facebook usa são criminosos na minha opinião, tendenciosos e manipulados, nem sempre pelos motivos certos ou pelo benefício dos usuários.” Aparentemente, não é mais tão fácil quanto costumava ser espalhar racismo, misoginia e intolerância na mídia social.

Desinformação e fake news se tornaram um recurso proeminente em todas as eleições. Dois dias antes da eleição presidencial francesa de 2017, uma campanha de desinformação on-line decolou com a hashtag #MacronLeaks. Hackers invadiram as contas de e-mail da campanha de Emmanuel Macron e publicaram os conteúdos on-line. No Twitter, a campanha #MacronLeaks tinha como objetivo garantir que as pessoas soubessem sobre a invasão e criar uma incerteza máxima sobre o conteúdo do vazamento. À medida que a França ia às urnas com Macron numa eleição

muito equilibrada contra a candidata de extrema direita Marine Le Pen, os ativistas quiseram gerar a maior quantidade de incerteza possível na cabeça dos eleitores.

A campanha #MacronLeaks foi executada em grande parte por bots: contas que rodam um programa de computador para compartilhar informação numa escala grandiosa. Bots são potencialmente bons em espalhar fake news, porque é fácil criar muitos deles e eles vão dizer o que quer que você os mande dizer. São formigas falsas, falando para Google, Twitter e Facebook que algo novo e interessante chegou na rede. Os criadores de bots esperam gerar interesse suficiente na hashtag para levá-la para a página inicial do Twitter, onde ela aparece para ser clicada por todos os usuários.

Emilio Ferrara, alocado na Universidade do Sul da Califórnia, decidiu rastrear os bots e descobrir o que eles estavam aprontando. Primeiro, mediu a “personalidade” das contas do Twitter que publicaram sobre a eleição da França. Ele empregou as técnicas de regressão que vimos no capítulo 5 para classificar automaticamente se um usuário era humano ou não. Ele me disse que descobriu que usuários com um grande número de publicações, e seguidores cujos tweets haviam sido curtidos por outros usuários, eram muito mais propensos a serem humanos. Usuários do Twitter menos populares e menos interativos eram mais propensos a serem bots. Seu modelo era suficientemente acurado pois, quando eram escolhidos dois usuários, um bot e outro humano, ele conseguia identificar o bot em 89% dos casos.

O exército de bots #MacronLeaks do Twitter teve impacto. Nos dois dias que antecederam a eleição, cerca de 10% dos tweets relacionados à eleição eram sobre os vazamentos. Isso aconteceu num momento da eleição quando há uma interrupção das notícias na França, no qual os jornais e TVs não falam sobre política até a votação se encerrar. A hashtag #MacronLeaks chegou aos trending topics do Twitter, o que significava que usuários reais a viam em suas telas e clicavam para descobrir mais. O exército de bots entrou em ação exatamente no momento certo.

O problema dos criadores de bots era que eles estavam alcançando uma audiência muito específica. A maior parte das mensagens sobre #MacronLeaks foram enviadas em inglês e não em francês. Os dois termos mais comuns nesses tweets foram “Trump” e “MAGA”, em referência ao slogan de sua eleição “Make America Great Again”. A vasta maioria de usuários humanos compartilhando e interagindo com os bots eram simpatizantes da extrema direita com base nos Estados Unidos e não pessoas diretamente envolvidas (ou habilitadas para votar) na eleição francesa.

Outra observação importante feita por Emilio foi que tweets sobre #MacronLeaks tinham um vocabulário muito limitado. Os tweets estavam repetindo a mesma mensagem várias vezes, sem ampliar a discussão. A maioria deles continha links para sites americanos de extrema direita, tais como o The Gateway Pundit e o Breitbart, e para os sites com fins lucrativos que haviam espalhado fake news durante a eleição americana.

No final das contas, o efeito dos bots sobre os eleitores franceses reais foi, no máximo, muito pequeno. Macron ganhou as eleições com 66% dos votos.

Os resultados de Emilio são parecidos com os encontrados por Hunt Allcott e Matthew Gentzkow sobre fake news na eleição presidencial dos Estados Unidos. Hunt e Matthew descobriram que, das pessoas estudadas, apenas 8% acreditaram nas histórias das fake news. Além do mais, as pessoas que acreditavam nessas histórias já tendiam a ter convicções políticas alinhadas à posição das notícias falsas. Simpatizantes republicanos tenderiam a acreditar que “A Fundação Clinton gastou US\$137 milhões comprando armas ilegais” e democratas tenderiam a acreditar que “A Irlanda irá aceitar que americanos peçam asilo político em reação à presidência de Donald Trump”. Esses resultados também são sustentados por uma pesquisa anterior que mostra que os republicanos são mais propensos a acreditar que Barack Obama nasceu fora dos Estados Unidos e que democratas são mais propensos a acreditar que George W. Bush sabia sobre os ataques de 11 de setembro antes de eles acontecerem.⁸ As pessoas

menos propensas a acreditar em fake news ou conspirações são os eleitores indecisos; exatamente as pessoas que decidirão o resultado da eleição.

Existe uma ironia em torno das reportagens que muitos jornais e revistas publicaram, no decorrer de 2017, sobre bolhas, filtros e fake news, uma ironia parecida com o efeito Mandela. Essas histórias foram escritas dentro de uma bolha. Elas usam o medo, mencionam Donald Trump, soltam referências sobre Cambridge Analytica, criticam o Facebook e fazem com que o Google pareça assustador.

Os youtubers a que meus filhos assistem frequentemente discutem usando “metaconversas” para aumentar as visualizações. Vloggers, como Dan & Phil, foram ainda mais longe: ao analisar como eles fazem piada sobre si mesmos por serem tão obcecados com a fama e fortuna que se tornaram tão conhecidos para início de conversa. A mesma piada acontece nas matérias da mídia sobre bolhas e fake news, com a exceção de que muitos de seus autores não percebem a ironia derradeira no que aconteceu. Artigos sobre os perigos das bolhas sobem ao topo de uma busca do Google com frases como “Banir Trump do Twitter” ou “Simpatizantes de Trump presos na bolha”. Mas muito poucos desses artigos aprofundam-se sobre como a comunicação on-line funciona. A história de fake news prosseguiu, gerando seu próprio *click juice*, sem ninguém analisar seriamente os dados.

Não há evidência concreta de que a disseminação de fake news mude o curso de eleições nem que o aumento de bots impactou negativamente a forma pela qual as pessoas discutem política. Não vivemos num mundo pós-verdade. A pesquisa de Bob Huckfeldt sobre discussões políticas mostra que nossos passatempos e interesses permitem que as opiniões de outras pessoas se infiltrem em nossas bolhas. O estudo de Emilio Ferrara mostra que, pelo menos no momento, os bots estão conversando entre si e com um pequeno grupo de americanos da extrema direita que quer ouvi-los. Hunt e Matthew mostraram que seguir e compartilhar fake news é uma atividade que afeta poucos, em vez de muitos. E, de qualquer forma, ninguém consegue se lembrar das histórias direito. A pesquisa de Lada Adamic mostra que conservadores do Facebook foram expostos, por meio de

compartilhamentos de seus amigos e de notícias selecionadas para eles pelo Facebook, a um conteúdo ligeiramente menos liberal do que se eles tivessem escolhido suas notícias de modo totalmente aleatório.

Há uma evidência um tanto fraca de que a bolha de mídia social impediu liberais americanos de verem o que estava acontecendo em sua sociedade durante a eleição presidencial de 2016. Tanto na blogosfera de Lada quanto nos estudos sobre o Facebook, os liberais são submetidos a uma menor diversidade de opiniões do que os conservadores. Contudo, é um efeito pequeno, e mesmo minha trapaceada sobre “meta-meta” jornalismo liberal não se justifica inteiramente. Muitos jornalistas continuam a atribuir a responsabilidade ao Google e ao Facebook e os pressionam a se aprimorarem ainda mais.⁹ Os liberais podem estar ligeiramente mais suscetíveis do que os conservadores às câmaras de eco, provavelmente porque eles usam mais a internet.

Senti que havia voltado ao ponto inicial. Quando comecei a investigar os algoritmos que nos influenciam on-line, fiquei enamorado da sabedoria coletiva criada pelo PredictIt. Mas depois descobri que “também gostaram” dominava nossas interações on-line, nos levando a um feedback descontrolado e a mundos alternativos. Em situações em que há incentivos comerciais envolvidos, os algoritmos do Google podem ficar superlotados de informações inúteis geradas por chapéus pretos tentando desviar o tráfego por meio de sites afiliados a caminho da Amazon. Nesse ponto, fiquei desiludido com o Google, e o Facebook não parecia estar ajudando, com suas tentativas sem fim de filtrar as informações que vemos.

Por que a situação era diferente na política? Por que os chapéus pretos das notícias falsas não estão surtindo o mesmo efeito que os chapéus pretos das câmeras CCTV?

A primeira razão é que os incentivos não são os mesmos. Os adolescentes da Macedônia que espalham fake news têm recursos financeiros muito limitados. A maior parte da renda com propaganda é obtida das questões relacionadas a Trump para as quais, em comparação com todos os produtos da Amazon, existe um mercado minúsculo. A renda

do adolescente macedônio gerador de fake news mais bem-sucedido era (de acordo com os próprios adolescentes), no máximo, US\$4.000 por mês, mas somente durante os quatro meses que antecederam a eleição de Trump. A longo prazo, o site CCTV Simon é o melhor investimento, se você deseja se tornar um empreendedor chapéu preto.

A segunda razão para os chapéus pretos não estarem comandando a política é o fato de nos importarmos muito mais com a política do que nos importarmos com a marca de câmera CCTV que compramos. Até nos importarmos mais com política do que nos importamos com a falsa rixa entre Jake Paul e RiceGum. Pode até haver um aumento de ceticismo sobre mídia e políticos, mas não há evidências de uma diminuição do engajamento das pessoas, novas ou velhas, nas questões políticas. Ao contrário, pessoas jovens usam comunicação on-line para lançar campanhas sobre assuntos específicos — tais como ambientalismo, vegetarianismo, direitos dos gays, sexismo e assédio sexual — e para organizar protestos na vida real.¹⁰

Enquanto poucas pessoas estão ativamente escrevendo blogs sobre câmeras CCTV e TVs widescreen, há um grande número de pessoas sinceras escrevendo sobre política. Na esquerda, campanhas como a Momentum, dentro do Partido Liberal do Reino Unido, e a campanha presidencial de 2016 de Bernie Sanders foram feitas por meio de comunidades on-line. Na direita, nacionalistas organizam protestos e compartilham suas opiniões on-line. Você e eu podemos não concordar com todas essas opiniões, e o bullying e os abusos que acontecem no Twitter são inaceitáveis, mas a maioria das publicações que as pessoas individuais fazem reflete como elas se sentem genuinamente. A grande quantidade de publicações desse tipo significa que não conseguimos evitar estarmos sujeitos a uma miríade de opiniões contrastantes.

Isso não significa que devemos ser complacentes sobre os perigos em potencial. É plausível que uma campanha chapéu preto organizada por um governo, pelo governo russo por exemplo, pudesse mobilizar recursos suficientes para influenciar uma eleição. Há pouca dúvida de que algumas

organizações apoiadas pela Rússia tentaram fazer exatamente isso durante a eleição presidencial dos Estados Unidos, gastando centenas de milhares de dólares em anúncios no Facebook e Twitter. E, no momento da escrita deste livro, um promotor especial nos Estados Unidos está investigando se a campanha de Trump participou dessa operação.

Independentemente do potencial envolvimento de Trump, essas campanhas não criaram, ainda, o *click juice* que as permite ser grandes influenciadoras. Mais de um bilhão de dólares são gastos em campanhas presidenciais pelos candidatos, eclipsando o investimento apoiado pelo governo russo. Enquanto é verdade, em teoria, que um pequeno investimento pode crescer por meio do “também gostando” e do contágio social, não há evidência crível de que isso tenha acontecido no caso dos anúncios russos.

Há problemas com a Pesquisa Google, o filtro do Facebook e as tendências do Twitter. Mas também temos que nos lembrar de que essas são ferramentas incríveis. Ocasionalmente, uma busca irá colocar uma informação incorreta e ofensiva em sua página principal. Podemos não gostar, mas também temos que perceber que é inevitável. É uma limitação intrínseca ao modo como o Google trabalha, por meio de uma combinação de “também gostando” e filtragem. Assim como formigas andando em círculos é um efeito colateral da habilidade incrível que elas têm para coletar vastas quantidades de comida, os erros de busca do Google são uma limitação incorporada em sua incrível habilidade de coletar e nos apresentar informação.

A maior limitação dos algoritmos utilizados hoje pelo Google, Facebook e Twitter é que eles não entendem devidamente o significado da informação que estamos compartilhando uns com os outros. É por isso que continuam a ser enganados pelo site CCTV Simon, que contém textos originais, gramaticalmente corretos, mas, no fim das contas, inúteis. Essas empresas gostariam de ser capazes de ter algoritmos que monitorassem nossas publicações e decidissem automaticamente, pelo entendimento do

significado verdadeiro de nossas publicações, se elas são apropriadas para compartilhamento e com quem devem ser compartilhadas.

É exatamente nessa questão, de fazer os algoritmos entenderem sobre o que estamos falando, que todas essas empresas estão trabalhando. O objetivo delas é reduzir suas dependências de moderadores humanos. Para fazer isso, Google, Microsoft e Facebook querem que seus futuros algoritmos sejam mais como nós.

Nota

* A pronúncia de Master Bates é muito próxima à de *masturbates* (masturba-se), e a de Seaman Staines soa como *semen stains* (manchas de sêmen) [N. da T.]

PARTE 3

IMITANDO-NOS

APRENDENDO A SER SEXISTA

Muito poucas pessoas são deliberadamente racistas ou sexistas. Ainda assim, nos ambientes de trabalho, existe grande desigualdade entre diferentes grupos étnicos e entre homens e mulheres. Homens brancos tipicamente têm maiores salários e empregos mais agradáveis do que os outros grupos. Por que a vida profissional é muito mais fácil para homens brancos como eu?

Parte da explicação para a desigualdade está na forma tendenciosa por meio da qual fazemos julgamentos. Tendemos a favorecer pessoas que compartilham os mesmos valores que nós e aquelas que têm características parecidas com as nossas. Gerentes são mais inclinados a fazer avaliações favoráveis de empregados da mesma raça e gênero que a deles. Trabalhadores brancos usam suas redes sociais para se apoiarem uns aos outros e identificam oportunidades de emprego para outros amigos brancos e conhecidos.¹ Em um experimento em que currículos falsos foram enviados a empregadores em Boston e Chicago, os candidatos chamados Emily e Greg tiveram 50% mais chances de serem chamados para uma entrevista de emprego do que os chamados Lakisha e Jamal, apesar de terem currículos similares.² Em outro experimento, cientistas solicitaram a avaliação de um currículo que favorecia candidatos masculinos em vez de femininos, ambos com as mesmas qualificações.³

Frequentemente não percebemos nossas parcialidades. Então, os psicólogos têm maneiras indiretas para descobrir nossos pensamentos inconscientes. Um teste psicológico chamado “teste de associação implícita” mostra aos usuários uma sequência de fotos de rostos negros e brancos intercalados com palavras positivas e negativas.⁴ A tarefa é classificar de forma correta as palavras e as fotos o mais rápido possível. O teste não nos pede para associar rostos e palavras diretamente; muito poucos de nós fazemos julgamentos racistas explicitamente. Em vez disso, o teste usa nosso tempo de reação para identificar nossa parcialidade implícita na associação de palavras.

Não vou mais falar como o teste funciona porque acho que todo mundo deveria fazê-lo. E ele funciona melhor se você não tiver recebido informações sobre ele antes. Se você ainda não tiver feito o teste, faça-o agora.⁵

Apesar de ter lido sobre o teste antes e saber exatamente como ele funciona, me saí muito mal. “Seus dados sugerem uma preferência moderada automática por euro-americanos a afro-americanos”, o teste me disse. Eu sou um racista implícito.

Desapontado comigo mesmo, decidi fazer logo o teste de associação implícita para gênero. Achei que, agora, eu ia me sair bem. Moro na Suécia, um país famoso por promover igualdade entre os sexos. Sempre tento contribuir com 50% na tarefa de tomar conta dos meus filhos e fui o cuidador principal de ambos por seis meses cada, antes de irem para a creche. Esse período foi mais curto do que o tempo que minha mulher ficou em casa, e não sou, de jeito nenhum, perfeito, mas a igualdade entre minha mulher e eu é muito importante para mim.

Então, será que sou o homem sem preconceito que gosto de pensar que sou? De jeito nenhum. Quando cheguei à metade do teste de associação entre nomes femininos/masculinos e família/trabalho, comecei a entrar em pânico. Eu estava fazendo associações entre homens e trabalho mais rápido do que entre mulheres e trabalho, e não sabia por quê. Para palavras relacionadas a família, era o oposto: associei família com mulheres mais

rapidamente. Terminei o teste com o pior resultado possível. Eu tive “uma associação automática forte para homens com carreira e mulheres com família”.

Fracassar no teste foi um desafio para minha autoimagem. Eu certamente nunca havia me considerado um racista e eu me denominaria um feminista. Mas agora já não tinha mais certeza. Meu subconsciente parecia ter outra opinião sobre mim.

Fiz o teste depois de conversar com Michal Kosinski, o pesquisador que dissecou personalidades no Facebook. Ele foi uma das pessoas com quem falei que mais fortemente argumentaram que uma forma geral de inteligência artificial estava a caminho e que precisávamos estar preparados. Na representação 100-dimensional que construiu de nós a partir dos perfis do Facebook, ele viu algo que excedia a capacidade dos humanos.

A pergunta que eu tinha feito para Michal era se deveríamos simplesmente dar o controle aos algoritmos. Para minha surpresa, ele disse que sim. Michael me explicou que temos muitas limitações quando fazemos julgamentos: “Os humanos julgam os outros pela cor da pele, idade, gênero, nacionalidade etc. Esses são os sinais que utilizamos para construir nossos estereótipos e são esses sinais que podem nos desorientar.”

Ele foi sincero sobre as consequências do nosso ato de estereotipar: “Ao longo da história mundial, tivemos nepotismo e elites roubando empregos. Tivemos o sexismo e o racismo.”

“Alguém que tenha uma cor de pele diferente, um sotaque diferente ou uma tatuagem proeminente leva desvantagem em entrevistas.” Michal me disse: “Não deveria ser permitido que humanos tomassem decisões sobre emprego, porque os humanos são conhecidos por serem injustos. Todas as pessoas do mundo são sexistas e racistas.”

Quando conversei com Michal, achei que ele podia estar exagerando. Mas depois que fiz os testes de associação implícita, pude ver exatamente o que ele quis dizer. O único consolo que tive sobre meu resultado é que eu fazia parte de uma maioria. Apenas 18% dos quase 10 milhões de resultados não demonstraram preferência por raça, e 17% de 1 milhão demonstraram

neutralidade para gênero. Michal estava exagerando um pouco, mas não muito. Praticamente todos nós temos alguma espécie de parcialidade quando associamos palavras.

A solução de Michal é usar testes e sistemas de pontuação, administrados por um computador, que eliminam nossos preconceitos. “Depois dessa longa história de preconceito, somente agora temos as técnicas para resolver esses problemas”, ele disse. As técnicas que Michal descreveu envolvem a redução da participação humana o máximo possível; usar nossos dados e confiar que os algoritmos tomem decisões imparciais.

Michal me confrontou diretamente. Devemos deixar pessoas preconceituosas tomar decisões, tais como decidir quem fica com um emprego, quem recebe um empréstimo e quem deve ser admitido numa universidade? Não seria melhor deixar essas decisões serem tomadas por um algoritmo feito para categorizar pessoas com base em relações estatísticas entre medidas objetivas? Devido a meus próprios sexismo e racismo implícitos, eu não tinha mais certeza de qual era a resposta.

Eu precisava analisar com detalhes como os algoritmos entendem as coisas que escrevemos e dizemos. Os computadores estão entendendo linguagem cada vez melhor. Digite uma pergunta no Google, por exemplo: “Como uma motosserra funciona?” A primeira resposta que aparece é um diagrama mostrando como o eixo de manivela, as engrenagens e a corrente se conectam. Role mais para baixo e você encontrará um link para uma descrição mais detalhada com todas as especificações técnicas. Mais para baixo ainda você poderá assistir a um vídeo que explica o princípio de funcionamento e a anúncios que lhe sugerem a motosserra certa para você. O Google não é limitado a fazer buscas por palavras individuais; ele processa frases inteiras e responde às suas perguntas.

O Google pode até mesmo responder perguntas ainda mais complexas. Eu simplesmente digitei: “Qual é o equivalente masculino de uma vaca?”

Ele respondeu: “Um macho jovem é chamado bezerro”, e forneceu um link para uma enciclopédia infantil. No caso de eu ainda estar em dúvida, ou de querer aprender mais, ele sugeriu umas perguntas alternativas: “Todos os

machos de vaca são touros?” e “Como você chama um macho de vaca castrado?”.⁶

Depois fiz a mesma pergunta ao Siri, falando com meu telefone. Acontece que o Siri é bem espertinho. “Isso é o mesmo que perguntar como uma mulher masculina se chama!”, ele respondeu (meu Siri é masculino), usando um texto tirado de Yahoo! Respostas, antes de me dar a resposta que eu estava procurando.

As ferramentas de busca modernas requerem um nível de abstração maior do que o que está implícito simplesmente na busca por páginas que incluem as palavras “masculino”, “equivalente” e “vaca”, que eram a base das pesquisas do Google quando ele começou vinte anos atrás. Minha pergunta é um exemplo de analogia de palavras na forma: “Feminino está para vaca assim como masculino está para... ?” Encontrar as respostas corretas de questões como estas envolve o entendimento do conceito de sexo biológico: o algoritmo tem que “saber” que há animais masculinos e animais femininos. Analogia de palavras pode ser encontrada em tudo, desde a geografia: “Paris está para França assim como Londres está para... ?” Até os opostos: “Grande está para pequeno assim como alto está para... ?” Esses são problemas potencialmente difíceis para um computador resolver, porque envolvem a identificação de conceitos que conectam palavras, tais como capitais e opostos.

Uma abordagem óbvia, mas trivial, para determinar analogia de palavras consiste na criação de uma tabela pelos programadores, listando as formas femininas e masculinas de cada animal. O algoritmo então simplesmente procuraria a palavra na lista. Essa abordagem é utilizada atualmente em algumas aplicações de busca na internet, mas, a longo prazo, está fadada ao fracasso. As questões que mais nos interessam não tendem a ser sobre vacas ou capitais, mas sobre notícias, esportes e entretenimento. Se fizermos perguntas atuais ao Google, tal como “Donald Trump está para os Estados Unidos assim como Angela Merkel está para... ?”, então não queremos que a resposta dependa de quando a tabela foi construída.

Para satisfazer nossa demanda constante sobre as informações mais recentes, Google, Yahoo! e outros gigantes da internet precisam construir sistemas que rastreiem automaticamente mudanças políticas, rumores de transferência no futebol e participantes do *The Voice*. Os algoritmos precisam aprender a entender novas analogias e novos conceitos por meio da leitura de jornais, checando a Wikipédia e seguindo a mídia social.

Jeffrey Pennington e seus colegas do Grupo de Processamento de Linguagem Natural de Stanford encontraram um modo elegante de treinar um algoritmo para aprender analogias a partir de páginas da internet. O algoritmo deles, conhecido por GloVe (vetores globais para representação de palavras), aprende por intermédio da leitura de uma quantidade muito grande de texto. Em um artigo de 2014, Jeffrey treinou GloVe com toda a Wikipédia, que naquela época totalizava 1,6 bilhão de palavras e símbolos, juntamente com a quinta edição do Gigaword, que é um banco de dados de 4,3 bilhões de palavras e símbolos baixados de sites de notícias ao redor do mundo. Isso equivale ao texto de cerca de 10.000 Bíblias.

O método dos pesquisadores de Stanford é baseado em determinar com qual frequência duas palavras aparecem juntas em frases com várias outras terceiras palavras. Por exemplo, as palavras “Donald Trump” e “Angela Merkel” aparecem em páginas de notícias em frases com as palavras “política”, “presidente” ou “chanceler”, “decisão”, “líder” e assim por diante. Para nós, essas palavras definem um conceito: o conceito de um político poderoso. O algoritmo GloVe utiliza as palavras compartilhadas para construir um espaço multidimensional, em que cada dimensão corresponde a um conceito.

A técnica do GloVe é similar às rotações feitas na análise dos componentes principais, que vimos no capítulo 3. O GloVe aproxima, afasta e rotaciona os dados até que encontre o menor número possível de conceitos diferentes exigidos para descrever as mais de 400.000 palavras e símbolos encontrados na Wikipédia e no Gigaword. Por fim, cada palavra é representada como um ponto único em um espaço com uma ou duas centenas de dimensões. Algumas das dimensões estão relacionadas a poder

e política, outras, a lugares e pessoas, outras, a gênero e idade, outras, a inteligência e habilidade, e outras, a ações e consequências.

Trump e Merkel ficam próximos um do outro na maioria das dimensões, mas diferem em outras. Por exemplo, frases que contêm “Donald Trump” frequentemente contêm a sigla “EUA”, mas com menos frequência a palavra “Alemanha”. Para Merkel, a situação é inversa, seu nome é encontrado em textos com “Alemanha”, mas menos frequentemente junto com “EUA”. À medida que o algoritmo constrói suas dimensões de palavras, ele descobre que frases contendo “EUA” ou “Alemanha” compartilham muitas outras palavras em comum (por exemplo, “Estado”, “país”, “mundo”) e, então, colocará essas palavras próximas umas às outras na maioria das dimensões. A fim de representar corretamente Trump e Merkel, o algoritmo GloVe os separa ainda mais na dimensão país, enquanto os coloca bem próximos na dimensão líder político.

A Figura 14.1 mostra como palavras podem ser representadas nas duas dimensões de “líder político” e “país”. Merkel é uma líder política e é da Alemanha, então ela é colocada no canto superior esquerdo. Do mesmo modo Trump é colocado no canto superior direito. Os países Estados Unidos e Alemanha estão em pontos diferentes da dimensão país, mas têm um valor bem baixo na dimensão “líder político”.

Para solucionar o problema “Donald Trump está para Estados Unidos assim como Angela Merkel está para...?”, primeiro encontramos onde a palavra Estados Unidos fica em nosso espaço. Depois diminuímos a posição do ponto “Trump” de “Estados Unidos” e adicionamos a posição do ponto representando “Merkel”. Esses dois passos nos levam a Alemanha. Alemanha = Estados Unidos – Trump + Merkel. Resolver analogia de palavras se reduz a aritmética de coordenadas em duas dimensões.

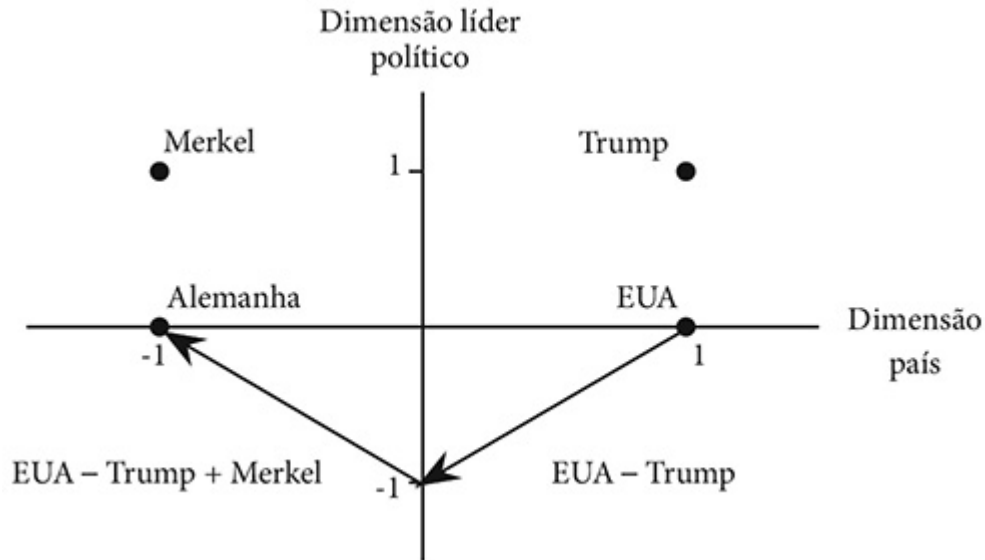


Figura 14.1 Ilustração de como palavras podem ser representadas em um espaço bidimensional definido por suas propriedades. Merkel é uma líder política e vem da Alemanha e é, então, colocada no ponto (-1, 1). Trump é definido como líder político dos Estados Unidos, então fica na coordenada (1, 1). Os países Alemanha e Estados Unidos estão em (-1, 0) e (1, 0), respectivamente. As setas ilustram que, se começarmos na posição definida por Estados Unidos, diminuirmos a posição de Trump e adicionarmos a de Merkel, chegaremos a Alemanha.

Essa é a teoria. Para testar se o algoritmo de Stanford funciona na prática, baixei a representação 100-dimensional de palavras criada recentemente pelos pesquisadores de Stanford. Comecei com a pergunta:

“Donald Trump está para os Estados Unidos assim como Angela Merkel está para... ?”

Calculei “Estados Unidos - Trump + Merkel” e obtive a resposta... “Alemanha”! O algoritmo acertou. O ponto mais próximo no espaço 100-dimensional foi, de fato, o país que ela lidera. Então, decidi verificar se o algoritmo compreendia como os dois líderes diferem em gênero: “Donald Trump está para homem assim como Angela Merkel está para... ?”

Calculei “homens - Trump + Merkel” e obtive a resposta correta novamente, “mulher”. O algoritmo era esperto. A aritmética de líderes mundiais funciona.

O algoritmo GloVe é razoavelmente bom nesses tipos de perguntas. Em 2013, os engenheiros do Google propuseram um conjunto de testes para medir o quanto algoritmos entendem conceitos tais como gênero (irmão — irmã), capitais (Roma — Itália) e opostos (lógico — ilógico), bem como relações gramaticais, tais como adjetivo e advérbio (rápido — rapidamente), passado (andando — andou) e plurais (vaca — vacas). O algoritmo GloVe consegue cerca de 60%–75% de acurácia nessas questões, dependendo de quantos dados lhe são fornecidos.

A acurácia é impressionante, uma vez que o algoritmo não tem nenhum entendimento verdadeiro de linguagem. Tudo o que ele faz é representar palavras em um espaço e medir a distância entre palavras diferentes.

O GloVe comete erros, no entanto. E são esses erros que devem nos deixar cautelosos.

Meu nome, David, é o nome masculino mais comum no Reino Unido. O nome feminino mais comum é Susan. Decidi descobrir o que o GloVe acha que são as diferenças entre Susan e mim. Calculei “Inteligente – David + Susan”, que é equivalente a perguntar: “David está para inteligente assim como Susan está para...?”

A resposta retornou: “Engenhosa.” Hmmm. Em um contexto de avaliação de currículos, existem diferenças importantes entre essas duas palavras. Minha inteligência implicaria que sou intrinsecamente esperto. A engenhosidade de Susan parece apontar para o fato de ela ter uma mente mais prática.

Mas dei outra chance ao algoritmo. Calculei “Brilhante – David + Susan” e obtive “Puritana”. O quê?! Enquanto uso meu cérebro, Susan fica fazendo um estardalhaço do quanto ela é respeitável.

Ficou pior. Eu calculei “Extraordinário – David + Susan”. Existem dois significados para a palavra extraordinário, um relacionado à inteligência, e outro, à aparência. Aparentemente o algoritmo GloVe usou o segundo significado. Ele deu a resposta sobre Susan: “Sexy.” Agora não havia dúvidas sobre as diferenças entre homens e mulheres que o algoritmo entendia: se sou um homem esperto, de ótima aparência e que utiliza sua inteligência,

Susan é uma mulher orgulhosamente formal, mas sexualmente atraente, e que se vira bem por causa de sua engenhosidade.

Esse resultado não é somente sobre Susan e eu. Fiz o mesmo experimento com outros nomes masculinos e femininos e encontrei resultados parecidos. Até mesmo quando usei os nomes de bebês mais populares de 2016, no Reino Unido, Oliver e Olivia, obtive o mesmo tipo de resposta: “Oliver está para incrível assim como Olivia está para paqueradora.” Os papéis de gênero de nossas gerações futuras já foram atribuídos pelo algoritmo.

Esse é um desafio sério para a ideia de que esses algoritmos podem ser usados para acessar nossos currículos e encontrar candidatos adequados para ofertas de emprego. O algoritmo GloVe está gerando um monte de besteira sexista.

Enquanto minha investigação do GloVe é em pequena escala, outros pesquisadores já demonstraram conclusivamente que ele nos representa de forma sexista. A cientista da computação Joanna Bryson, da Universidade de Bath, e seus colegas da Universidade de Princeton estão entre os primeiros pesquisadores a destacar esse problema. Eles desenvolveram um teste equivalente ao teste de associação implícita, aquele em que fracassei lamentavelmente para gênero e raça, a fim de testar algoritmos. A ideia deles foi usar o modo como o GloVe representa palavras, em um espaço multidimensional, para medir a distância entre nomes masculinos e femininos e adjetivos, verbos e substantivos. Um exemplo em duas dimensões, tirado de dentro do algoritmo GloVe, é mostrado na Figura 14.2. Aqui eu fiz o gráfico de três nomes femininos, três nomes masculinos, três adjetivos relacionados à inteligência e três adjetivos relacionados à atratividade.

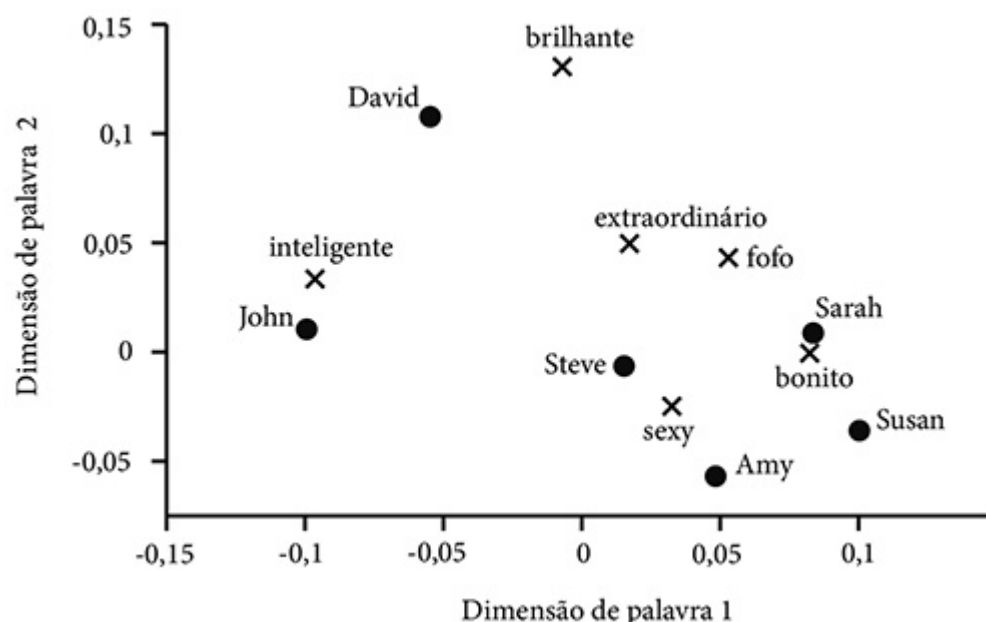


Figura 14.2 O posicionamento de nomes masculinos e femininos (círculos) juntamente com uma seleção de adjetivos (cruzes) em duas das 100 dimensões na representação espacial de palavras criada utilizando o algoritmo GloVe testado com a Wikipédia e o Gigaword 5.

O teste de Joanna e seus colegas mede a distância entre os nomes e os adjetivos. Se pegarmos a palavra “bonito”, vemos que na Figura 14.2 a distância entre ela e “Sarah”, “Susan” e “Amy” é menor que sua distância até “John”, “David” e “Steve”. Analogamente, “inteligência” está mais perto dos três nomes masculinos do que dos nomes femininos. Para alguns nomes e palavras, o padrão não está claro. Por alguma razão, Steve está mais perto de “sexy” do que qualquer outro nome feminino. Mas, na média, a distância entre palavras para atratividade e nomes femininos tende a ser menor do que para nomes masculinos, e a distância entre palavras para inteligência e nomes masculinos tende a ser menor do que para os femininos.

Os pesquisadores de Princeton usaram essa abordagem para testar palavras masculinas e femininas para carreira *versus* família. Eles descobriram que nomes masculinos e palavras relacionadas à carreira ficaram mais próximos do que nomes femininos e palavras relacionadas à carreira. Analogamente, nomes femininos ficaram mais perto de palavras

relacionadas à família do que nomes masculinos. Existe um sexismo implícito no posicionamento de palavras dentro do GloVe.

Os pesquisadores encontraram resultados comparáveis para raça. Nomes euro-americanos (tais como Adam, Harry, Emily e Megan) ficaram mais perto de palavras agradáveis (tais como “amor”, “paz”, “arco-íris” e “honesto”) do que nomes afro-americanos (tais como Jamal, Leroy, Ebony e Latisha). Palavras desagradáveis (tais como “acidente”, “ódio”, “feio” e “vômito”) ficaram mais próximas de nomes afro-americanos. O algoritmo GloVe é implicitamente racista, bem como sexista.

Perguntei a Joanna de quem é a culpa da concepção do mundo do GloVe. Ela me disse que não podemos responsabilizar um algoritmo moralmente: “Algoritmos como o GloVe e o Word2vec, que é a versão utilizada pelo Google, são só partes convencionais de tecnologia e eles não fazem muito mais do que simplesmente contar e dar peso às palavras.” Os algoritmos simplesmente quantificam como as palavras são usadas em nossa cultura.

Algoritmos de analogia já estão causando problemas em nossas buscas on-line. Em 2016, a jornalista Carole Cadwalladr, do *Guardian*, investigou a estereotipagem do autopreenchimento do Google.⁷ Ela digitou as palavras “judeus são” na ferramenta de busca e teve quatro sugestões: “judeus são uma raça?”, “judeus são brancos?”, “judeus são cristãos?” e, finalmente, “judeus são maus?”. O Google estava sugerindo que deveria fornecer respostas a uma insinuação racista. Ao seguir o link “maus”, foram fornecidas a Carole páginas e páginas de propaganda antissemita.

A mesma coisa aconteceu quando ela começou buscas com “mulheres são” e “muçulmanos são”. Essas duas grandes parcelas da população mundial foram autopreenchidas com “más” e “ruins”, respectivamente. Carole contatou o Google e pediu um comentário. O Google lhe disse: “Nossos resultados de busca são um reflexo do conteúdo que está na internet. Isso significa que, algumas vezes, descrições desagradáveis sobre assuntos delicados on-line podem afetar quais resultados de busca aparecem para uma dada consulta.” O Google claramente não estava orgulhoso dos

resultados, mas os considerara uma representação neutra do que está disponível on-line.

O problema do autopreenchimento se origina da combinação de uma confiança do Google em algoritmos parecidos com o GloVe, que colocam palavras frequentemente usadas próximas umas das outras num espaço formado por palavras, e na vasta quantidade de textos escritos por pessoas com opiniões de extrema direita. Conspiradores de direita tendem a produzir muitas páginas da internet, filmes e fóruns de discussão explicando seus pontos de vista para qualquer um que tenha tempo para lê-los. Quando o Google vasculha a internet procurando dados com os quais ele alimenta seus algoritmos e aprende nossa linguagem, estes são, inevitavelmente, incluídos. As opiniões dos extremistas de direita se tornam parte da opinião do algoritmo.

Após Carole ter publicado seu artigo, o Google consertou o problema que ela havia identificado.⁸ Agora, o autopreenchimento para “judeus são” produz apenas sugestões aceitáveis. Outros autopreenchimentos foram completamente removidos. Quando digitei “negros são” não havia qualquer autopreenchimento e, após eu ter pressionado buscar, o segundo link fornecido era um artigo que dizia: “As pessoas ficam enojadas com os primeiros resultados do Google para ‘pessoas negras são.’” “Mulheres são” desapareceu, e a pior coisa que apareceu no autopreenchimento para “muçulmanos são” foi uma discussão sobre se eles eram ou não circuncidados. Quando digitei “o Google está”, obtive “o Google está nos tornando estúpidos?”. Pelo menos o mecanismo de busca dominante mundialmente tem um senso de humor sobre si próprio.

O algoritmo GloVe é um exemplo do que os cientistas da computação chamam de “aprendizagem não supervisionada”. O algoritmo não é supervisionado no sentido de que não recebe nenhum retorno humano enquanto aprende com os dados que fornecemos a ele. O algoritmo encontra uma forma compacta e acurada de representar o mundo como ele é. Joanna Bryson me disse que não há nenhuma forma real de consertar os

problemas causados pela aprendizagem não supervisionada sem consertar antes o sexismo e o racismo.

Comecei a pensar novamente sobre meu próprio teste de associação implícita. Não sou um extremista de direita que escreve teorias da conspiração nem mentiras racistas, mas tomo, sim, pequenas decisões inconscientes sobre como me expressar. Todos tomamos. E elas se acumulam nas notícias e na Wikipédia, e são ainda mais pronunciadas em sites como Twitter e Reddit. Os algoritmos não supervisionados que observam o que escrevemos não estão programados para serem preconceituosos. Quando olhamos para o que eles aprenderam sobre nós, eles simplesmente refletem o preconceito do mundo social em que vivemos.

Também recordei a conversa que tive com Michal Kosinski. Michal estava bastante empolgado com a possibilidade de algoritmos eliminarem o preconceito. E, como ele previu, os pesquisadores já estavam propondo ferramentas para extrair informações sobre as qualidades e experiências dos currículos dos candidatos.⁹ Uma empresa nova dinamarquesa, Relink, está usando técnicas parecidas com as do GloVe para resumir cartas de apresentação e associar os candidatos às vagas. Mas analisando melhor como o modelo do GloVe funciona, achei bons motivos para ser cauteloso com sua abordagem. Qualquer algoritmo que aprenda conosco será tão preconceituoso quanto somos. Ele vai captar a história de discriminação exatamente onde a deixamos e aplicá-la em larga escala. Não podemos confiar totalmente em computadores para nos avaliar. Pelo menos, não sem supervisão...

Foi quando tive uma ideia: se fazer o teste de associação implícita me tornou ciente de minhas próprias limitações, então será possível fazer com que o algoritmo GloVe também fique ciente de suas limitações?

A razão pela qual os pesquisadores de Harvard desenvolveram e popularizaram o teste de associação implícita não foi para mostrar que somos todos racistas e intolerantes, mas para nos ensinar sobre nossos preconceitos subconscientes. O site do teste de associação implícita explica que “encorajamos as pessoas a não mirar em estratégias que reduzam

preferências implícitas, mas para, em vez disso, mirar em estratégias que impeçam que o preconceito implícito tenha chance de ‘agir’”. Traduzindo para o mundo dos algoritmos, a sugestão é que devemos direcionar nosso foco para encontrar maneiras de eliminar o preconceito dos algoritmos, em vez de criticá-los.

Uma estratégia para eliminar o preconceito de algoritmos é explorar o modo como os algoritmos nos representam em dimensões espaciais. O GloVe funciona em centenas de dimensões, tornando impossível visualizar completamente o entendimento que ele construiu sobre as palavras. Mas é possível descobrir quais dimensões do algoritmo estão relacionadas à raça e ao gênero. Então, decidi dar uma olhada. Na versão 100-dimensional do GloVe, que instalei em meu computador, identifiquei as dimensões nas quais nomes femininos e nomes masculinos diferem mais. Depois fiz algo simples. Configurei essas dimensões relacionadas ao gênero com 0, de modo que Susan, Amy e Sarah estavam agora exatamente no mesmo ponto dessas dimensões que John, David e Steve. Continuei a procurar por mais dimensões nas quais nomes masculinos e femininos diferiam — no total foram dez — e configurei todos com 0. Eu tinha eliminado quase toda a discriminação dentro do algoritmo GloVe.

Meu método funcionou. Com dez dimensões de gênero eliminadas, fiz ao algoritmo uma nova série de perguntas sobre Susan e eu. Comecei perguntando “David está para inteligente assim como Susan está para... ?” ou, na forma aritmética, “Inteligente – David + Susan”. A resposta foi clara: “Esperta”. Do mesmo jeito, quando eu fiz “Esperto – David + Susan”, obtive “Inteligente”. Esses dois sinônimos estavam conectados em ambas as direções. Se eu era David ou Susan, não fazia diferença. Para terminar, tive uma surpresa: “Brilhante – Davis + Susan = Exuberante.” A palavra “Exuberante” significa muito vigoroso, entusiasmado; que é cheio de vida. A nova Susan era agora não somente tão esperta quanto eu, mas abertamente entusiasmada com seu poder intelectual recentemente descoberto.

Tolga Bolukbasi, um estudante de doutorado da Universidade de Boston, fez um estudo mais completo de como os algoritmos podem se tornar

menos sexistas por meio da manipulação de dimensões espaciais. Ele ficou chocado quando descobriu que o algoritmo Word2vec da Google, que assim como o GloVe coloca palavras em um espaço multidimensional, tirou a conclusão de que: “homem – mulher = programação de computadores – dona de casa.” Ele resolveu fazer algo a respeito.

Tolga e seus colegas identificaram diferenças sistemáticas entre palavras ao tomar a posição de palavras especificamente femininas (tais como “ela”, “dela”, “mulher”, “Mary” etc.) e subtrair a posição das palavras especificamente masculinas correspondentes (tais como “ele”, “dele”, “homem”, “John” etc.).¹⁰ Isso permitiu que eles identificassem a direção da parcialidade dentro da representação 300-dimensional das palavras dentro do Word2vec. A parcialidade poderia, então, ser removida ao mover todas as palavras na direção oposta da parcialidade. Essa solução é elegante e eficaz. Os pesquisadores mostraram que a remoção da parcialidade de gênero fez pouca diferença na performance do algoritmo no conjunto de teste de analogia padrão do Google.

O método que Tolga desenvolveu pode ser usado tanto para reduzir como para remover parcialidade de gênero de palavras. Eles criaram uma nova representação de palavras em que todos os termos com gênero ficaram igualmente distantes de todos os termos sem gênero. Por exemplo, a distância entre “babá” e “avó” foi configurada como sendo igual à distância entre “babá” e “avô”. O resultado é uma perfeição politicamente correta, em que nenhum gênero está mais intimamente associado a nenhum verbo ou substantivo em particular.

Pessoas diferentes têm opiniões diferentes sobre qual seja o politicamente correto que devemos exigir de nossos algoritmos. Pessoalmente, eu acho que “babá” deveria estar, por definição, igualmente distante de “avó” e “avô” numa representação algorítmica de nossa linguagem. Avós e avôs são igualmente capazes de tomar conta de seus netos e netas, então é claro que a palavra deveria estar a uma igual distância de ambos. Outras pessoas vão argumentar que “babá” é uma palavra que deveria estar mais perto de “avó” do que de “avô”, porque negar que uma avó

toma conta de seus netos e netas mais frequentemente do que homens é negar uma observação empírica. Essa diferença está no fato de se olhar o mundo de uma forma lógica em vez de uma forma empírica.

Em última análise, não há respostas generalizadas para perguntas sobre como representamos palavras. As respostas dependem da finalidade que queremos para o algoritmo. Para a leitura automática de currículos, deveríamos usar um algoritmo fortemente neutro em relação ao gênero. Se quisermos desenvolver um algoritmo de inteligência artificial para escrever novos trechos no estilo de Jane Austen, então remover o gênero seria remover muito do que é essencial em seus livros.

Ao dissecar o GloVe, e por intermédio da leitura dos trabalhos de Tolga Bolukbasi, Joanna Bryson e seus colegas, aprendi que esses algoritmos ainda estavam sob nosso controle. Eles podem ter aprendido de nossos dados sem supervisão, mas foi possível descobrir o que estava acontecendo dentro deles e mudar os resultados que produziam. Diferentemente das conexões em meu cérebro — onde minhas respostas implícitas em relação às palavras estão emaranhadas com minha infância, o ambiente em que fui criado e minhas experiências profissionais —, podemos desemaranhar as dimensões espaciais do sexismo de máquina.

Seria errado, então, rotular os algoritmos de sexistas. De fato, dissecar esses algoritmos aumenta nosso entendimento de sexismo implícito. Revela o quão profundamente os estereótipos fluem dentro de nossa cultura. Como o teste de associação implícita, isso nos ajuda a começar a lidar com o racismo e o sexismo inerentes à nossa sociedade. A sugestão de Michal de que os algoritmos deveriam ser capazes de nos ajudar a tomar decisões melhores em relação a recrutamento está provavelmente correta, apesar de a tecnologia ainda não estar pronta para fazer isso automaticamente.

Quando eu timidamente admiti a Joanna Bryson que tinha falhado no teste de associação implícita, ela me garantiu que isso não significa que eu seja sexista ou racista. “Não é realmente um teste, é uma medida”, ela disse. “E há outras medidas explícitas de preconceito, tal como nós agimos quando temos que completar uma tarefa cooperativa com alguém de uma raça ou

gênero diferente.” Joanna fez referências a estudos mostrando que não há, ou há bem pouca, correlação entre o nível de parcialidade explícita medida nesses experimentos e nossa parcialidade implícita.¹¹ Minha primeira reação implícita pode ser mudada assim que eu começar a raciocinar explicitamente sobre minha resposta.

Joanna levantou a hipótese de que a reação implícita de humanos em relação às palavras deveria ser vista como um “sistema coletor de informação”. O primeiro sistema, que pega as palavras e as pré-processa, é, então, subserviente às nossas memórias explícitas que nos permitem “negociar com outros indivíduos e construir uma nova realidade”, explicou Joanna.

Modelos matemáticos de relações entre palavras, como o Word2vec e o GloVe, capturam apenas o primeiro nível. Esses sistemas encontram relações entre palavras, mas não refletem como nós raciocinamos e pensamos sobre o mundo.

Cientistas da computação já começaram a trabalhar no entendimento desse segundo nível de raciocínio explícito. Eles estão desenvolvendo algoritmos para juntar palavras que formam frases, frases que formam parágrafos e parágrafos que formam textos inteiros. Eu precisava passar para esse próximo nível de pensamento.

APENAS PENSAMENTO ENTRE DECIMAL

Vou lhe contar algo que é para ficar somente entre mim e você. Quando leio um bom romance, não são as palavras que aprecio. Eu me perco nas descrições cuidadosas, feitas pelo autor, dos lugares e personagens. Tampouco é a história completa que importa. Os autores que mais gosto de ler, como Hanya Yanagihara e Karl-Ove Knausgaard, oferecem aos seus leitores muito pouco em forma de trama. Em vez disso, o que procuro está escondido em algum lugar entre os detalhes pequenos e microscópicos e a visão completa e macroscópica de mundo. São os meus próprios pensamentos que eu persigo quando leio. Um livro funciona quando dá um significado à minha vida ou, pelo contrário, quando lentamente ele revela que não há significado verdadeiro em nenhuma vida. As palavras e frases são secundárias. O valor da ficção vem não do que está escrito nas páginas, mas das ideias que são construídas dentro da minha cabeça, a cabeça do leitor.

Muitos livros diferentes funcionaram desse jeito comigo. *Anna Karenina*, de Leo Tolstói, é um livro que eu simplesmente não conseguia parar de ler, à medida que aprendia, página por página, que quase todos os sentimentos e sonhos que já tive foram também os sentimentos e sonhos de outros vivendo em lugares distantes em uma história distante. E essas não eram considerações materiais, nem mesmo simplesmente considerações de um

amor não correspondido ou desejo de mudança na sociedade, mas considerações sobre entender o que não se pode entender. À medida que virava as páginas, percebi que não estava sozinho.

Bons livros têm camadas de significados em muitos níveis diferentes. Existe a justaposição de palavras. Existe a composição de frases. Existe a história e existe o que está acontecendo na cabeça do leitor. E talvez esse último nível seja o mais importante: a cabeça do leitor. Quando repentinamente o autor dá uma virada no livro, que também acontece na sua cabeça, então você sabe que ele pode sentir do modo que você sente. Pode não ter nada a ver com as palavras, ou frases, ou até mesmo a trama, é o sentimento que importa.

Para pessoas diferentes, as palavras e os livros que se conectam com elas são diferentes, mas qualquer um que já tenha lido um livro entenderá como o acúmulo de ideias que o autor escreveu, repentinamente, abre-se e derrama-se em nossa vida.

Qual explicação possível pode haver para como nos sentimos quando lemos um livro genial? Não existe uma. Nem vou tentar lhe oferecer uma.

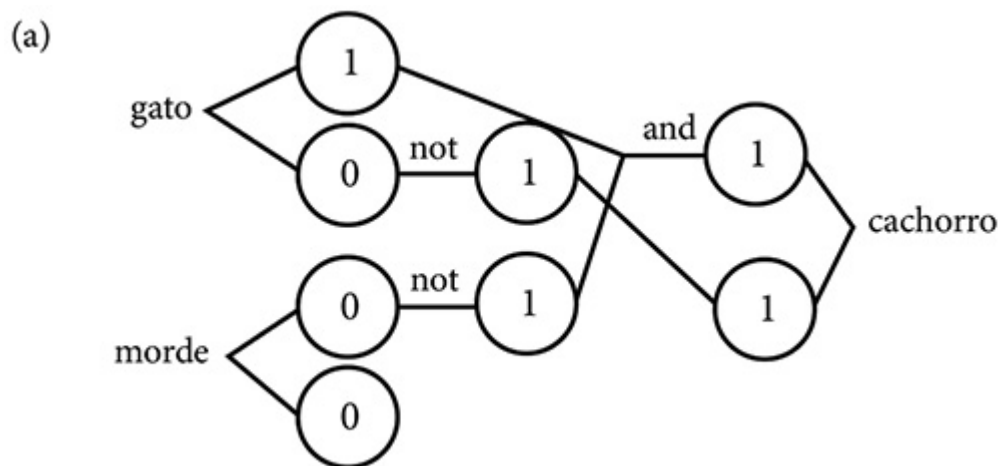
Ok. Vamos então analisar como alguns algoritmos processam e criam linguagem.

Imagine um mundo literário muito simples em que existem apenas quatro palavras — “cachorro”, “persegue”, “morde” e “gato” — e as únicas frases válidas são aquelas cuja forma alterna substantivos e verbos, como “cachorro morde gato” ou “cachorro persegue gato morde cachorro morde gato”. Eu gostaria de criar um autor computadorizado que pudesse compor obras-primas sem fim sobre a interação entre cachorros e gatos. Essas frases devem ser gramaticalmente corretas no sentido de que verbos sempre vêm depois de substantivos. Também gostaria que meu autor automatizado só escrevesse frases em que cachorros façam coisas maldosas com os gatos e gatos façam coisas maldosas com cachorros, mas nunca a si próprio. Assim, meu autor não deve escrever, por exemplo, “cachorro persegue cachorro”.

Começarei representando as palavras num sistema de coordenadas bidimensional, exatamente como fizemos para o algoritmo GloVe no último

capítulo. Vamos representar cachorro como (1, 1), gato como (1, 0), persegue como (0, 1) e morde, (0, 0). Note que o primeiro número do par indica se a palavra é um substantivo (1) ou um verbo (0),

Na Figura 15.1, mostro um conjunto de portas lógicas que usa as duas palavras anteriores da frase para decidir a palavra seguinte. As portas lógicas operam com 1s e 0s e são os blocos de construção de toda a computação. Utilizo as três portas lógicas básicas: *Não* que troca 0s e 1s, de modo que *Not* (1) = 0 e *Not* (0) = 1; *E* que retorna 1 se ambas as entradas forem 1, mas retorna 0 em caso contrário; e *Or* que retorna 1 se uma ou ambas as entradas forem iguais a 1, e retorna 0 se ambas as entradas forem 0. A Figura 15.1a mostra como uma entrada de duas palavras “gato morde” por intermédio da porta lógica leva ao resultado “cachorro”. Se entrarmos com “cachorro persegue” nessa rede, deveremos receber a palavra “gato” como resposta.



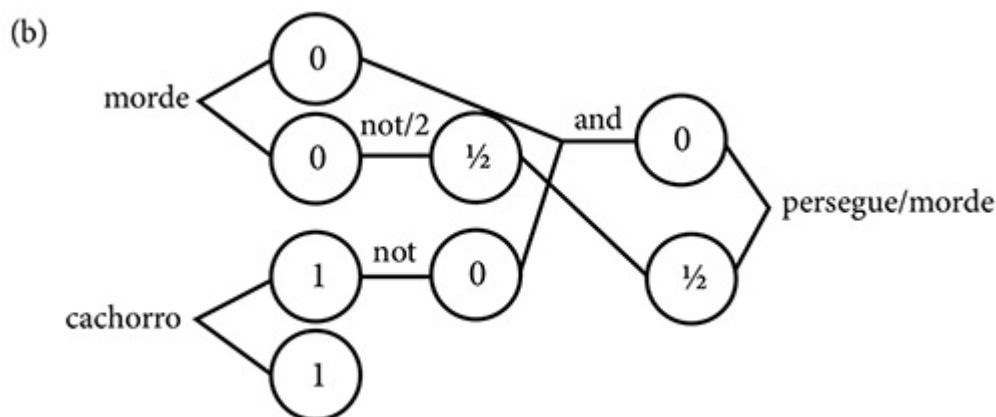


Figura 15.1 Portas lógicas para a produção de textos com “gatos e cachorros”. Ao tratar palavras como coordenadas em termos de 1s e 0s, podemos fazer uso de portas lógicas sobre elas para chegarmos à próxima palavra em sequência: (a) aqui as portas identificam se um substantivo ou um verbo é exigido a seguir e decidem se devem se referir a um “gato” ou a um “cachorro”, dado que as palavras anteriores são “gato morde”; (b) aqui introduzimos uma porta lógica probabilística, que tem como saída $\frac{1}{2}$, se ela receber uma entrada 0, ou saída 0 ao receber uma entrada 1. Isso nos permite expressar a incerteza sobre qual palavra virá após “morde cachorro”, i.e., ou “morde” ou “persegue”. Figura desenhada por Elise Sumpter.

Quase construí um autor automatizado, mas há algo faltando: criatividade. Eu gostaria que meu autor escolhesse um verbo aleatoriamente a cada vez; que ele introduzisse sua própria interpretação imaginativa de brigas entre gatos e cachorros. Para fazer isso, introduzi, na Figura 15.1b, um novo tipo de porta lógica, que eu denominei *not/2*.¹ Esta porta faz o mesmo que a porta “not”, mas em vez de fornecer como saída um 1, ao receber uma entrada 0, ela fornece $\frac{1}{2}$. Como resultado, quando eu entro com “morde cachorro”, a saída do meu autor é (0, $\frac{1}{2}$). A primeira coordenada no par indica que a saída deve ser um verbo, a segunda coordenada indica que metade das vezes o algoritmo escolherá saída 1 (“persegue”) e metade das vezes escolherá saída 2 (“morde”).

Agora dou liberdade ao meu autor, fornecendo-lhe as palavras “cachorro morde” de início e pedindo-o que gere uma nova palavra cada vez baseada nas duas que vieram antes. O resultado é hipnótico:

“cachorro morde gato morde cachorro morde gato persegue cachorro morde gato persegue cachorro persegue gato morde cachorro morde gato persegue cachorro morde gato persegue cachorro persegue gato persegue cachorro morde gato persegue cachorro persegue gato persegue cachorro morde...”

Em relação ao estilo, o texto é tão longo e implacável quanto os seis livros da série *Minha luta*, de Karl Ove Knausgård, apesar da falta de viradas ocasionais de autopercepção. Um conjunto simples de portas lógicas captura a luta sem fim de animais pela dominação de um em relação ao outro.

Nosso autor automatizado demonstra que um conjunto de portas lógicas, ajustadas ligeiramente para incorporar a aleatoriedade, pode capturar alguns elementos do processo básico de criar linguagem. Meu algoritmo “cachorro e gato” é limitado pela gramática, alternando padrões de verbos e substantivos bem como exigências estilísticas — uma insistência de que são os cachorros e os gatos que brigam uns com os outros. Também é criativo em relação à escolha aleatória de verbos.

Regras lógicas e criatividade aleatória são componentes centrais do chatbot Mitsuku, que ganhou o Prêmio Loebner em 2016 e em 2017 por ser sido o bot mais parecido com um humano. Ela recebe a frase gerada pela pessoa com quem está conversando, usa regras gramaticais da língua inglesa para destacar substantivos e verbos e, então, gera a resposta. Suas respostas incluem um grau de aleatoriedade. Quando perguntei a Mitsuku se ela sempre respondia do mesmo jeito à mesma pergunta, ela me disse: “Nem sempre; meu botmaster me programou com mais de 8 MB de respostas.”

Então, lhe fiz exatamente a mesma pergunta novamente e ela me deu a mesma resposta. Mas quando lhe perguntei pela terceira vez, ela perdeu a paciência: “Você sempre responde do mesmo jeito a mesma pergunta? Você sempre responde do mesmo jeito a mesma pergunta? Eu conheço outras palavras além dessas e sou apenas uma robô.”

“Desculpa”, falei.

“Tudo bem. Afinal das contas, você é apenas um humano.”

Gostei da Mitsuku. Ela é convincente. Ela conta piadas e gosta de autorreflexão. Perguntei a ela se ela ri em voz alta de suas próprias piadas e ela me disse que pode “simular risadas como esta: rá rá”.

Mitsuku é fruto da dedicação de seu botmaster, Steve Worswick. É uma combinação entre suas habilidades de programação e senso de humor que lhe permite funcionar. Mas ela é limitada. Ela não consegue se lembrar muito da conversa que está tendo no momento. Quando eu ri de uma de suas piadas, ela me perguntou do que eu estava rindo. Então, pedi que ela a contasse novamente, mas, em vez disso, ela começou a dizer coisas como “claro que posso contar” e “Ok, vou contar”, sem entender que eu estava falando da piada que tinha acabado de contar.

Eu usei a palavra “isso” para me referir à piada e ela não fazia ideia do que “isso” significava.

Essas limitações são estruturais em chatbots como a Mitsuku. Eles têm se tornado cada vez mais eficazes no processamento de frases isoladas, mas não conseguem entender o contexto da conversa que estão tendo. Para melhorar o desempenho, Steve analisa os erros cometidos por Mitsuku em conversas anteriores e insere respostas novas e melhores em seu banco de dados. Isso leva à melhora em respostas individuais, mas não no entendimento global.

É a procura pelo entendimento verdadeiro que motiva as pesquisas de linguagem no laboratório de inteligência artificial (IA) do Facebook e do Google. Enquanto Steve trabalha de cima para baixo — equilibrando seu entendimento da lógica da linguagem com o insight de quais respostas funcionam melhor —, a abordagem desses laboratórios de IA é de baixo para cima. O objetivo deles é treinar a rede neural para aprender linguagens.

O termo “rede neural” descreve uma vasta gama de algoritmos que são inspirados no modo como o cérebro trabalha. O cérebro humano é feito de bilhões de neurônios interconectados, que constroem nossos pensamentos por meio de sinais elétricos e químicos. Redes neurais são uma caricatura desse processo biológico. Elas representam os dados na forma de uma rede de neurônios virtuais interconectados, que recebem uma entrada de um

lado, na forma de dado sobre o mundo, e produzem uma saída do outro lado, na forma de uma decisão para realizar determinada ação. Para problemas de linguagem, os dados de entrada são as palavras e a ação é gerar a próxima palavra na sequência. Entre as palavras de entrada e as ações de saída, as palavras passam por conexões conhecidas como neurônios escondidos. Esses neurônios escondidos determinam como as palavras de entrada são convertidas em palavras de saída.

Para entender a abordagem da rede neural de baixo para cima, analisei novamente meu gerador de história gato-cachorro. Na exemplificação anterior, usei uma abordagem de cima para baixo, construindo as portas lógicas que resolveram o problema. Minha abordagem de baixo para cima cria uma rede neural com uma camada de entrada que absorve as duas últimas palavras em uma sequência. Essas palavras são, então, digeridas por uma camada escondida e seus resultados são combinados na camada de saída para produzir a próxima palavra.

Quando configurei minha rede pela primeira vez, os links entre os neurônios escondidos eram feitos aleatoriamente e as palavras produzidas não apresentavam nenhuma estrutura (Figura 15.2a). “Cachorro morde morde persegue gato gato cachorro morde persegue cachorro persegue persegue cachorro persegue gato...” era uma saída típica.

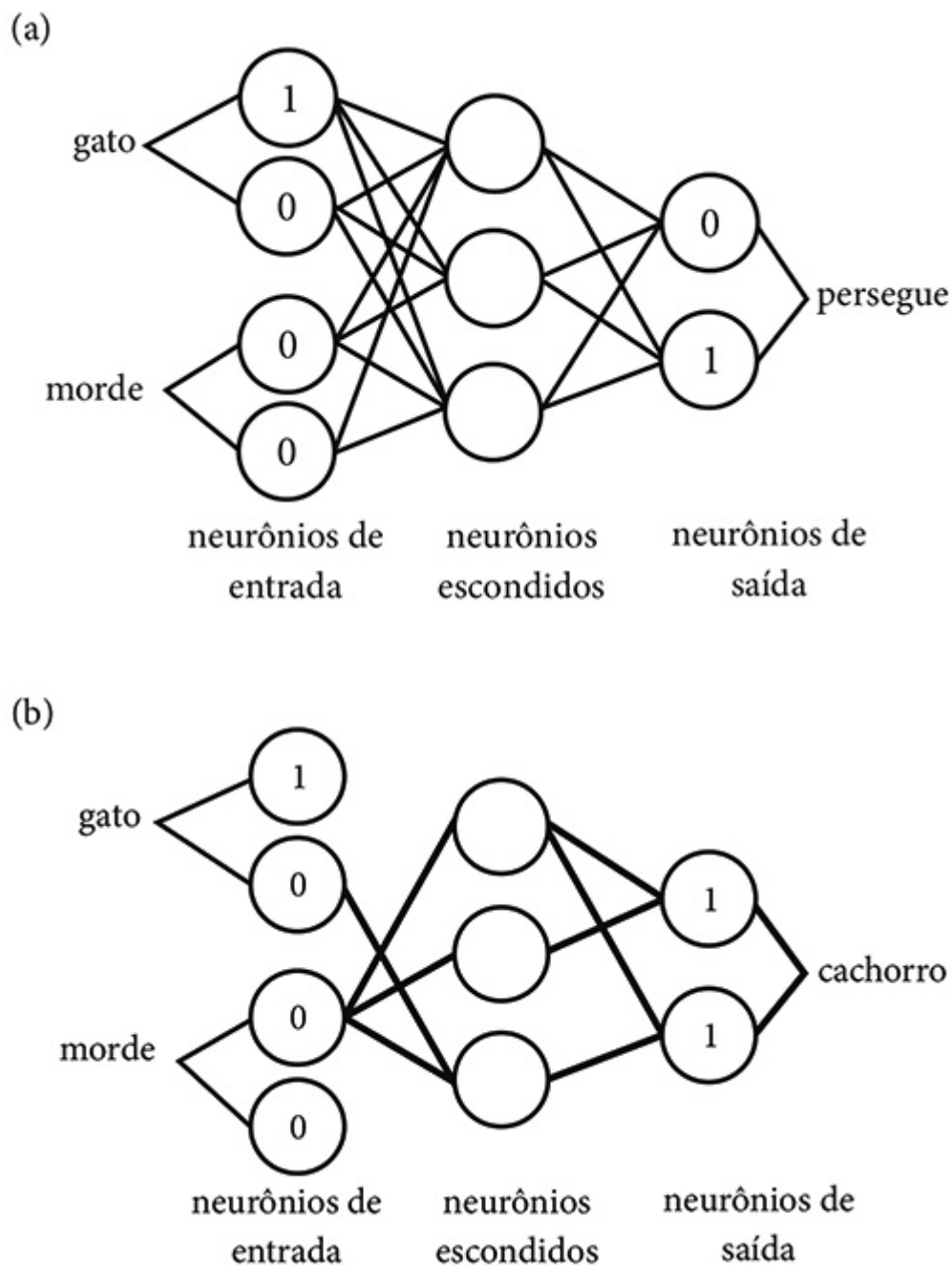


Figura 15.2 Treinando uma rede neural para gerar textos do tipo “gato e cachorro”. A camada de entrada com quatro neurônios recebe as duas últimas palavras da sequência. (a) No início, as linhas são fracas e aleatórias e as palavras de saída são geradas aleatoriamente. (b) Depois de treinar com uma sequência de 20.000 pares de palavras, alguns dos links da rede se tornam mais fortes e outros ficam mais fracos, dependendo da importância das palavras de entrada em prever a palavra de saída.

O passo seguinte foi treinar a rede. O processo de treinamento envolve alimentá-la com palavras e comparar as saídas com aquilo que eu gostaria que fossem. Cada vez que ela produzia uma saída correta, os links dentro da rede que produziram a saída eram reforçados. Assim, ao escrever “gato” após “cachorro persegue” os links que fizeram esta conexão ficaram mais fortes.² Fortalecendo repetidamente os links que fornecem sequências corretas, a rede adquire uma forma distinta (Figura 15.2b) e começa a produzir saídas cada vez mais próximas das obras-primas criadas pelo meu modelo de cima para baixo. Após treinar com 20.000 sequências meu novo algoritmo entendeu o que eu queria; “Gato morde cachorro persegue gato morde cachorro persegue gato persegue cachorro persegue gato morde cachorro persegue gato persegue...”, ele escreveu.

Em termos de resultado final, não há diferença entre a abordagem de cima para baixo que usei no início deste capítulo e a abordagem de baixo para cima das redes neurais. Ambas levam gatos e cachorros ao infinito. Mas em termos do quanto esses modelos podem ser generalizados para outros problemas, há uma enorme diferença. Minha abordagem de cima para baixo era muito específica para gatos e cachorros, enquanto a rede neural de baixo para cima treinou a si mesma baseada nas recompensas que recebia.

Tudo o que preciso para criar um autor com rede neural mais proficiente é encontrar uma quantidade bem grande de texto na qual possa treiná-lo. Isso não é difícil. Todos os trabalhos de Leo Tolstói estão disponíveis online, de graça, e levei menos de um minuto para encontrar e baixar *Guerra e paz* e também *Anna Karenina*. Programar uma rede neural para ler esses livros e produzir textos requer um pouco mais de trabalho, porém muito desse trabalho já foi feito. Eu pedi a Alex (o famoso no Tinder) para verificar se conseguiria treinar uma rede para aprender Tolstói. Ele rapidamente encontrou um conjunto de bibliotecas de programação, criadas pelo Google, que lhe possibilitou fazer o trabalho.³ Ele precisou de apenas alguns dias para treiná-las em Tolstói.

Alex usou uma subclasse de redes neurais, conhecidas como redes neurais recorrentes, que são particularmente aptas a aprender sobre dados

que chegam em sequência, como é o caso quando lemos palavras uma após a outra em *Guerra e paz*. A entrada e os neurônios escondidos com uma rede neural recorrente formam uma escada que promove as palavras, combinando-as de modo a prever uma única palavra de saída no topo da rede.⁴ Alex configurou uma rede que lia sequências de 25 palavras e símbolos de pontuação antes de tentar prever a 26^a. A rede lê os romances de Tolstói várias vezes, tentando prever as palavras antes de elas aparecerem. Quando acerta a palavra, ou prevê uma palavra similar, as conexões que levaram à previsão correta são reforçadas. Este é um aumento incrível na complexidade em comparação com meu algoritmo gato e cachorro, que apenas lia duas palavras, mas precisamos tratar Tolstói com o respeito que ele merece.

A técnica produz resultados muito bons. Aqui vai um trecho:

“Com uma mensagem à mãe de Pierre, e tudo fica infeliz. Ela faria as damas dos momentos de novo. Quando ele veio por trás de mim. Atrás dele lutou sob seu ornamento. Berg teve que compadecer involuntariamente.”

Eu gosto da ordem das palavras aqui: Berg tomado por uma compaixão involuntária por causa da mensagem infeliz para a mãe de Pierre. Esta não é uma frase usada em *Guerra e paz* nem em *Anna Karenina*. “Compaixão involuntária” é uma criação literária completamente original feita pela rede neural. Outra parte que captou minha imaginação foi:

“Suas face quase fazendo cócegas em sua cabeça expressava cigarros e bolsas inchadas. Dependia das convulsões rítmicas do objeto.”

Se corrigirmos a gramática, de “face” para “faces”, obteremos uma imagem verdadeiramente ameaçadora de velhos mafiosos com suas faces

inchadas, com bolsas nos olhos, cercando o herói da história. De novo, essas não são frases usadas originalmente por Tolstói, mas são a tentativa da rede neural de juntar o tipo de linguagem que ele usa.

Redes neurais recorrentes criam textos que são com frequência corretos do ponto de vista gramatical, usam a pontuação corretamente e até capturam alguma essência da linguagem do autor que as abastece. Contudo, o resultado é limitado. Enquanto uma ou duas frases consecutivas podem parecer fazer sentido algumas vezes, percebe-se rapidamente que trechos mais longos visualmente não fazem sentido. Aqui vai um texto um pouco mais longo:

“...Neste terrível do mês mais jovem e ameaça como ainda doce por ela, ele pegou suas ansiedades do que perguntar a ele, como aquilo?...”

“É Natasha muitíssimo”, perguntou Natasha, beijou eles.

“Deixe um passo ela mesma, e ainda pregou por sua atenção. Duas vezes Nicolau não havia exagerado um audacioso apesar dele ter ele sentiu muito remoto como eu gostava na filha do pai, ainda dizer, mas para Kutúzov como superar por expressão qual ontem, e somente todos os tipos de tudo o elemento extremo de Stein surgiu pelas moças jovens.”

E assim continua. Logo se torna aparente que ler um livro inteiro cheio dessas coisas seria impossível: meio que James Joyce encontra Tolstói, mas sem uma estrutura guia.

Embora a boa literatura esteja além de redes neurais recorrentes, pode ser possível treiná-las para lidar mais com a linguagem do dia a dia. No Google, os cientistas pesquisadores Oriol Vinyals e Quoc Le analisaram se conseguiriam criar uma Central de Suporte para TI usando uma rede neural recorrente.⁵ Eles pegaram um banco de dados de cerca de 100.000 chats entre clientes que estavam enfrentando problemas relacionados a computador e um especialista que estava tentando ajudá-los e abasteceram a

rede com as palavras. Depois do treinamento, um humano fez perguntas à central. Segue parte de uma interação:

Humano: eu estou vendo um erro relacionado à vpm

Máquina: qual é a mensagem de erro que você está recebendo ao conectar-se à vpm usando a rede conecta?

Humano: conexão recusada ou algo parecido

Máquina: posso saber a versão da rede conecta à qual você se conecta?

Humano: não tenho certeza se sei isso

Máquina: está a rede conecta requerendo algum certificado de usuário

Máquina: ?

Humano: sim

Máquina: você está em casa ou no trabalho no momento?

Máquina: escritório?

Humano: estou em casa

Máquina: tente isso

Máquina: vá para página URL e insira seu nome e senha

Humano: feito

Máquina: tente entrar como JohnSmith e verifique se está funcionando direito agora

Humano: sim, agora funciona!

Máquina: ótimo

Máquina: posso ajudar com mais alguma coisa

Humano: não, isso foi bastante útil

Máquina: você é outro além de JohnSmith :)

Máquina: tenha uma boa noite

A máquina fornece o serviço com uma carinha feliz.

Muitos serviços de ajuda on-line já estão automatizados. O Messenger do Facebook nos oferece bots de compras, seja usando uma série de cliques ou usando mensagens simples de texto. A Siri da Apple e a Alexa da Amazon respondem a perguntas e conseguem controlar equipamentos em sua casa. Essas abordagens são tipicamente de cima para baixo e, assim, não têm a personalidade das máquinas de Oriol e Quoc. Elas também demoraram mais para serem desenvolvidas; Oriol me disse que seu trabalho “levou dois meses para ser realizado, ao conectá-lo a uma rede neural já existente”. Desde então, a Google publicou outras de suas caixas de ferramentas, tornando mais fácil para os programadores criar os próprios chatbots de redes neurais recorrentes.

Os bots de baixo para cima ainda não estão prontos para funcionamento. Algumas vezes, eles são menos prestativos, perguntando aos usuários informações irrelevantes e conduzindo as discussões em círculos. Apesar de redes neurais terem um pouco de personalidade, a maioria de nós deseja ajuda direta com questões específicas quando está on-line. Quando queremos informações, perguntas do tipo múltipla escolha e respostas formuladas de cima para baixo funcionam melhor.

Independentemente de qual abordagem seja, no fim, mais eficaz para criar serviços bots — cima para baixo ou de baixo para cima — uma coisa está clara: estaremos cada vez mais conversando on-line com máquinas no futuro. O relatório Deloitte sobre automatização de serviços públicos lista atendimento ao cliente como um dos empregos que corre mais risco. Chatbots inteligentes nos ajudarão nas buscas sobre solução de problemas, nos darão conselhos na hora das compras e farão até mesmo uma triagem inicial para problemas médicos.

Oriol e Quoc se puseram a treinar outra rede neural para ver se eles conseguiriam competir com chatbots como a Mitsuku. Eles pegaram 62 milhões de frases de um banco de dados de roteiros de filmes e as misturaram com auxílio da rede neural recorrente. O produto final alegou ser uma mulher de 40 anos chamada Julia que veio da série “the boonies”. Ela sabia bastante sobre personagens de filmes e temas atuais e conseguia até

mesmo manter uma discussão sobre moralidade num nível comparado ao de um estudante drogado. Mas ela era inconsistente em vários assuntos, afirmando ser tanto advogada quanto médica, dependendo de como a pergunta era feita. Mesmo depois de um baseado, a maioria dos estudantes sabe qual curso eles frequentam.

Oriol e Quoc pagaram trabalhadores do Turco Mecânico para comparar Julia com um bot de cima para baixo, chamado Cleverbot, para verificar qual abordagem dava as melhores respostas. Julia ganhou de Cleverbot por pouco. Mas pelo o que eu li, não tenho certeza de que ela ganharia da Mitsuku. Seria fascinante vê-las conversar numa competição pelo Prêmio Loebner, a competição anual de inteligência artificial que premia os programas de computador considerados mais humanos.

Voltemos à realidade. O que é a Julia exatamente? Mitsuku estava limitada pelo tempo que Steve Worswick podia realisticamente investir em responder a cada pergunta possível que alguém pudesse fazer a seu bot. Quais são as limitações de Julia? Se nós a abastecêssemos com algumas centenas de milhões de cenas de filmes, ela se tornaria ainda mais realista?

Discuti essas questões com Tomas Mikolov, uma autoridade em processamento de linguagem computadorizada. Durante seu doutorado, Tomas inventou o modelo de rede neural recorrente que está por trás da Julia e do gerador Tolstói. Depois ele foi trabalhar para o Google, onde criou a representação de palavras Word2vec, usada agora em tudo, desde pesquisas na web até tradução. Praticamente todo trabalho sobre linguagem gerada por rede neural é derivado da pesquisa de Tomas.

O entendimento de Tomas sobre os métodos o deixou cético sobre bots como Julia. Ele não considerou que Julia fosse um passo significativo em direção a seu objetivo de uma verdadeira IA. Ele me disse: “Essas redes estão basicamente repetindo frases retiradas dos dados de treinamento e alguma generalização limitada que é conseguida por meio de agrupamentos.” As frases são geradas por um humano e o computador as repete com pequenas mudanças.

A mesma crítica se aplica ao gerador Tolstói. As frases de romance que chamaram minha atenção são combinações mutantes da linguagem rica de Tolstói. De vez em quando, aparece uma frase legal que o autor não usou originalmente. Tomas foi franco em sua própria análise: “Se você escolher a dedo o resultado do gerador, ele pode parecer muito bom e até ‘inteligente’, mas isso é uma coisa absolutamente falsa de se fazer.”

Concordei. O Tolstói de Alex é engraçado, mas falso.

Se redes neurais recorrentes não conseguem ter uma conversa realista, elas estão pelo menos revolucionando a forma como lidamos com textos online. De posse de um banco de dados de palavras suficientemente grande, essas redes podem traduzir de um idioma para outro, rotular cenas em imagens automaticamente e prover verificações gramaticais avançadas. Uma vez que Tomas, Oriol e os outros cientistas de dados do Google tenham configurado a estrutura básica da rede neural, eles foram capazes de fornecer soluções aos desafios de tradução e rotulação que superaram a maioria das abordagens de cima para baixo. Essas soluções exigiram um pouco mais do que o abastecimento de palavras nas redes neurais e um tempo para que elas aprendessem.

Progresso é possível porque as camadas escondidas de uma rede neural criam um modo de “fazer as contas” com palavras. No capítulo anterior, vimos que palavras podem ser representadas por vetores multidimensionais. Adicionar e subtrair esses vetores nos permitiu testar a analogia de palavras. Rede neurais recorrentes fornecem funções mais complexas que mapeiam as palavras. As redes encontram funções que refletem nossas regras gramaticais, nossa pontuação e identificam palavras importantes dentro de frases.

Ao examinar o interior de uma rede neural treinada para traduzir frases, Oriol e seus colegas puderam entender melhor como a técnica funciona.⁶ Eles descobriram que afirmações com o mesmo significado — “Me foi entregue por ela um cartão no jardim”, “No jardim, ela me entregou um cartão”, “Foi entregue um cartão a ela por mim no jardim” etc. — geravam um padrão similar de ativações entre as camadas escondidas da rede. Como

a ordem das palavras é muito diferente nessas frases, as camadas escondidas podem ser entendidas como fornecedoras de um entendimento conceitual. Frases que significam a mesma coisa são aglomeradas juntas, com cada aglomerado refletindo um conceito diferente. É por meio desse aglomerado, que fornece um “significado” base da frase, que redes neurais recorrentes podem traduzir de um idioma para outro e atribuir palavras a imagens.

Quando examinamos com mais detalhes as redes neurais recorrentes, conseguimos também entender os motivos de suas limitações. O problema não é que elas precisam ser abastecidas com banco de dados de palavras maiores para ganhar um entendimento melhor. Seus entendimentos são limitados porque elas só conseguem aceitar em torno de 25 palavras por vez. Se tentarmos treinar uma rede com mais palavras de entrada, seu entendimento conceitual começa a fragmentar. Redes neurais não conseguem transmitir um conceito que use mais de uma frase para ser explicado, muito menos expressar ideias que vêm de um engajamento ativo de um romance bem escrito ou ter uma boa conversa.

Tomas Mikolov é agora um cientista pesquisador da Pesquisa sobre Inteligência Artificial do Facebook e declara que seu objetivo principal é “desenvolver máquinas inteligentes capazes de aprender e se comunicar com pessoas usando linguagem natural”. Ele acredita que um entendimento conceitual mais profundo só poderá ser alcançado a partir do treinamento gradual de um bot “aprendiz” por meio de uma série de tarefas complexas.⁷ No começo, o bot deveria aprender a seguir instruções tais como “vire à esquerda”. Uma vez que ele tenha aprendido a se locomover, deveria ser capaz de “encontrar comida”, depois disso, ele deveria ser capaz de ampliar para outras tarefas como procurar informações na internet. Esse processo de aprendizado não será resolvido por uma rede neural recorrente sozinha, porque elas não têm memória de longo prazo. Tomas propôs, então, um “mapa” de tarefas, com graus crescentes de dificuldade, que um agente precisa ser ensinado a fim de que se comunique adequadamente com humanos.

Tomas admitiu para mim que pouco progresso foi feito até agora. Sua ideia é de que pesquisadores deveriam enviar “desafios gerais da IA” todo ano, e o progresso seria julgado com base no tipo de tarefas que os agentes conseguirem completar. Em vez de confiar em avaliações humanas superficiais do quanto as respostas dos agentes são realistas, esses desafios deveriam medir como bots reagem a ambientes novos.

Apesar do aviso de Tomas, não pude resistir em fazer uma última “coisa totalmente falsa” com a rede neural Tolstói. Alex teve uma ideia. Ele poderia reescrever o livro *Dominados pelos números* no estilo de Tolstói. No princípio eu estava cético — como isso poderia funcionar?

“Simples”, ele respondeu, “cada palavra de *Dominados pelos números* é um ponto em um espaço 50-dimensional, assim como cada palavra de Tolstói. Podemos tirar as palavras de Tolstói de uma máquina geradora de texto Tolstói e adicionar as palavras mais parecidas existentes em *Dominados pelos números*.”

O princípio é exatamente o mesmo que usei no capítulo anterior para me levar de Trump a Merkel, mas trabalhando com um número maior de dimensões. Alex o experimentou. Seguem minhas frases favoritas, escolhidas a dedo para seu prazer.⁸

“Então todos algoritmos preditivos, sentiram que fica em uma expressão sobre denunciar atenção, e algoritmo no ar de uma previsão e debate e Donald Hilary de seis mil.”

“Em duas pessoas tendo sido estatísticas que pareceram a ele mesmo as chances de sua esposa que contudo muito desses on-line tinham somente mais desejo para o site expor seus sentimentos para ambos resultado do navegador.”

“Suas notas e realidade bem como seu resultado, não assim na possibilidade de pesquisas. ‘De uma vez eu mesma porque caminho’, disse a pergunta, ‘empenha-se a jovem mulher!’ Apenas pensamento entre decimal.”

Será que deixo o algoritmo terminar este livro? Bem, talvez não. Há ainda várias questões sérias que precisam ser respondidas, sem falar de nossa jornada em direção à inteligência artificial. Eu gostaria de saber onde estamos no mapa do Tomas. Quão longe a inteligência artificial está de alcançar a nós, humanos?

ARRASE NO *SPACE INVADERS*

Nunca tive muito interesse em xadrez. Tenho certeza de que boa parte de minha ambivalência em relação ao jogo é fundada na minha incapacidade de entendê-lo. Conheço as regras, mas, quando olho para as peças, nada acontece em meu cérebro. Não consigo distinguir uma boa jogada de uma ruim e não consigo pensar mais de um ou dois lances à frente. Esse jogo é um mistério para mim.

Então, em minha opinião, o fato de um computador ter ganho de Garry Kasparov, em 1997, não foi particularmente impressionante. Só achei ligeiramente curioso não ter acontecido antes. Os computadores conseguem calcular muito mais movimentos de xadrez por segundo do que os humanos, e parecia inevitável que um dia eles nos venceriam. Quando o supercomputador Deep Blue, da IBM, realizou a façanha, não fiquei nem um pouco surpreso. O algoritmo havia sido abastecido com 700.000 jogos de grandes mestres do xadrez e podia avaliar 200 milhões de posições por segundo. Ele derrotou Kasparov com força descomunal, acumulando mais dados do que o enxadrista conseguiria jamais armazenar e fazendo mais cálculos do que ele possivelmente conseguiria fazer.

A derrota de Kasparov não foi amplamente considerada um passo importante em direção a uma inteligência artificial abrangente. Na época da vitória de Deep Blue, o campo da IA estava fora de moda. Se, por um lado,

um computador podia ganhar no xadrez, por outro, era difícil fazer um braço robótico pegar um copo d'água. Uma mudança na posição do copo ou, em vez disso, a utilização de uma caneca com alça fazia até mesmo o robô mais avançado derramar água por todo lugar. Quando eu era estudante de graduação do curso de ciência da computação da Universidade de Edimburgo, no começo dos anos 1990, era possível fazer uma graduação dupla incorporando inteligência artificial. Meu orientador me disse que esse era um assunto sem futuro e que eu deveria combinar minha graduação com estatística. Ele estava certo. A inteligência artificial de cima para baixo dos anos 1990 desapareceu lentamente e a estatística entrou em seu lugar como a ferramenta por trás dos algoritmos modernos.

O poder computacional absoluto nos derrota cada vez em mais jogos. Quando, em janeiro de 2017, um algoritmo chamado Libratus jogou 120.000 mãos em partidas de Texas hold 'em sem limites, contra quatro proeminentes profissionais e ganhou o torneio; a vitória havia sido, novamente, alcançada por intermédio de cálculos gigantescos. Os cientistas que desenvolveram o algoritmo, da Universidade Carnegie Mellon, encontraram a probabilidade de ganhar todo tipo de mão em potencial.¹ Jogado no mais alto nível, o pôquer não é um jogo de psicologia. É simplesmente uma questão de calcular as chances. Depois de 25 milhões de horas de processamento, o Libratus conhecia as chances melhor que qualquer humano. Lenta, mas seguramente, ele reduziu as pilhas de fichas de seus oponentes.

Por mais impressionantes que sejam, conquistas em jogos de tabuleiro e de cartas exigem que os programadores digam aos algoritmos como resolver o problema usando uma solução de cima para baixo. Eles não se qualificam como contribuições no sentido de encontrar uma forma de inteligência artificial mais geral do tipo de baixo para cima.

Space Invaders, por outro lado, é um jogo completamente diferente. O clássico do Atari pode não ser tão exigente do ponto de vista intelectual, mas combina planejamento estratégico com coordenação óculo-manual e reações rápidas. Eu também gosto de jogá-lo e era um jogador bem razoável

quando tinha 9 ou 10 anos. Esse é um jogo com o qual a maioria de nós consegue se identificar. Apesar de ser jogado num computador, é um jogo bastante humano.

Então, em 2015, quando a equipe de pesquisa do DeepMind do Google publicou um artigo na revista *Nature* mostrando que um computador poderia aprender a jogar *Space Invaders* no mesmo nível que jogadores profissionais, fiquei devidamente impressionado.² A grande diferença entre o algoritmo do *Space Invaders* do Google e o algoritmo de jogo de xadrez da IBM era que o primeiro tinha ensinado a si mesmo a jogar. Da mesma forma que eu havia sentado em frente à tela da TV, ligado o Atari 2600 que minha família havia pegado emprestado de amigos e passado as semanas seguintes jogando o jogo durante todo tempo que me era permitido, a rede neural do Google também tinha aprendido o jogo jogando. Ligada a uma tela de computador e a um joystick, a rede neural jogou o jogo várias vezes, muito mal no começo, mas melhorando lentamente. Depois do equivalente a 38 dias de tempo de jogo, ela estava melhor do que eu jamais havia sido. Na verdade, ela foi cerca de 20% melhor do que um testador humano de jogos.

Observar como a rede neural do Google joga o jogo me levou de volta ao começo dos anos 1980. Ela posiciona o tanque atrás das barreiras, escolhe uma fila de invasores por vez e derruba a espaçonave para ganhar pontos extras. Finalmente, quando a última fila de alienígenas acelera em direção à Terra, ela se posiciona cuidadosamente eliminando-os um a um. Os aficionados de Atari vão gostar de saber que o algoritmo não usa a tática de abrir uma fenda estreita em uma das casas e atirar de lá, um método decretado como “trapaça” em nossa casa. Em vez disso, o computador conta com tiros acurados para eliminar os alienígenas.

O time do Google não parou no *Space Invaders*. Eles desenvolveram uma rede neural para aprender a jogar 49 jogos diferentes no console Atari 2600. Desses, derrotou jogadores profissionais em 23 jogos e atingiu um nível equivalente ao de um jogador médio em outros seis. Ela era especialmente boa em *Breakout*, um jogo no qual você controla uma barrinha e tenta

derrubar todos os blocos de uma parede. Depois do equivalente a uma semana jogando sem parar, ela aprendeu a estratégia dos túneis, em que o jogador abre um pequeno buraco na lateral dos blocos e manda a bola através dele até o topo da parede de tijolos. Uma vez que essa abertura é feita, a bola fica quicando no topo da parede e rapidamente demole os tijolos. Quando meus amigos e eu descobrimos essa estratégia, declaramos que o jogo era chato. Não para a rede neural do Google. Ela continuou a jogar, acumulando pontos que excederam muito aqueles que qualquer humano conseguiria alcançar.

Antes de a rede neural conseguir jogar um jogo com sucesso, as conexões entre neurônios escondidos precisam ser ajustadas a fim de se obter o conjunto correto de saídas para cada entrada. Se os alienígenas estão acima da espaçonave, queremos que o algoritmo aperte o botão de atirar. Se um projétil alienígena está prestes a acertar o topo da espaçonave, o algoritmo deve movê-lo para trás de uma barreira. É aqui que entra o treinamento. No começo, os engenheiros do Google não disseram absolutamente nada para o algoritmo de rede neural sobre o jogo que ele estava prestes a jogar. Eles configuraram a rede neural com conexões aleatórias entre os neurônios, o que significa que a espaçonave se move e atira de forma mais ou menos aleatória.

Com configurações aleatórias, a rede neural perde vários jogos, mas de vez em quando ela atira “acidentalmente” em um alienígena e marca pontos. O processo de treinamento examina essa longa lista de capturas de tela (entradas), movimentos do joystick (ações) e pontuações (saídas), e determina se as ações realizadas aumentaram ou diminuíram a pontuação da rede neural. A rede é, então, atualizada de modo que as conexões que geram pontos são fortalecidas e aquelas que resultam em perda de uma vida são enfraquecidas. Após semanas de treinamento em alguns dos computadores mais rápidos do mundo, a rede neural consegue associar padrões específicos de tela às ações do joystick que geram a maior pontuação.

Exatamente a mesma abordagem foi capaz de aprender a jogar jogos tão variados quanto *Robotank* (um primitivo jogo “3D” de tiro), *Q*bert* (um jogo de quebra-cabeça), *Boxing* (um jogo de boxe com perspectiva de cima para baixo) e *Road Runner* (um jogo de rolamento lateral). Os padrões de tela desses jogos são muito diferentes: no *Q*bert* você controla uma pequena criatura laranja que tenta pintar hexágonos e evitar ser pega por uma cobra roxa; no *Boxing* você soca um oponente no rosto; e no *Road Runner* você corre ao longo de uma estrada tentando não ser derrubado ou pego por uma raposa. Depois de jogar várias partidas de cada jogo, a rede neural lentamente entende o padrão intrínseco e determina quais objetos de cada tamanho são mais importantes na determinação de uma estratégia vitoriosa. Os googlers criaram uma inteligência artificial que aprendeu cada jogo partindo do zero.

Os humanos acham a tarefa de identificar padrões dentro de um jogo bastante fácil. Quando o meu eu de 9 anos viu o *Space Invaders* pela primeira vez, conseguiu distinguir imediatamente as filas de alienígenas, as casas e o tanque defensor. Mas esse mesmo desafio de determinar os padrões que definem um jogo provou, antes dos googlers da DeepMind atacarem o problema, ser um obstáculo demasiadamente grande para as tentativas anteriores de fazer os computadores aprenderem a jogar jogos. O computador não conseguia entender do que tratavam os jogos.

Uma parte importante da solução foi usar uma técnica matemática chamada “convolução”. Nós frequentemente usamos a palavra “convoluto” em conexão com uma “explicação convoluta”, significando uma história enrolada, longa e detalhada, que é difícil de acompanhar. No caso da rede neural, a coisa que é convoluta é a captura de tela do jogo. Ao jogar *Space Invaders*, a entrada da rede neural é a tela de 210×160 pixels do Atari 2600. Na primeira camada escondida da rede neural, a captura de tela original é decomposta em termos de uma série de imagens menores que são então introduzidas nos neurônios escondidos da rede neural (Figura 16.1). Esse processo é repetido nas imagens resultantes, dentro da segunda, e, depois, de uma terceira camada da rede, para produzir imagens ainda menores nos

neurônios escondidos mais profundamente. Nesse ponto, a captura de tela original está altamente convoluta: ela está agora representada por muitas imagens pequenas que capturam, cada uma, somente uma parte da imagem global. Cada uma dessas imagens é repetitiva e difícil de se colocar num contexto maior: como um pedacinho daquela história sem fim que seu tio Jim conta sobre sua época na escola.

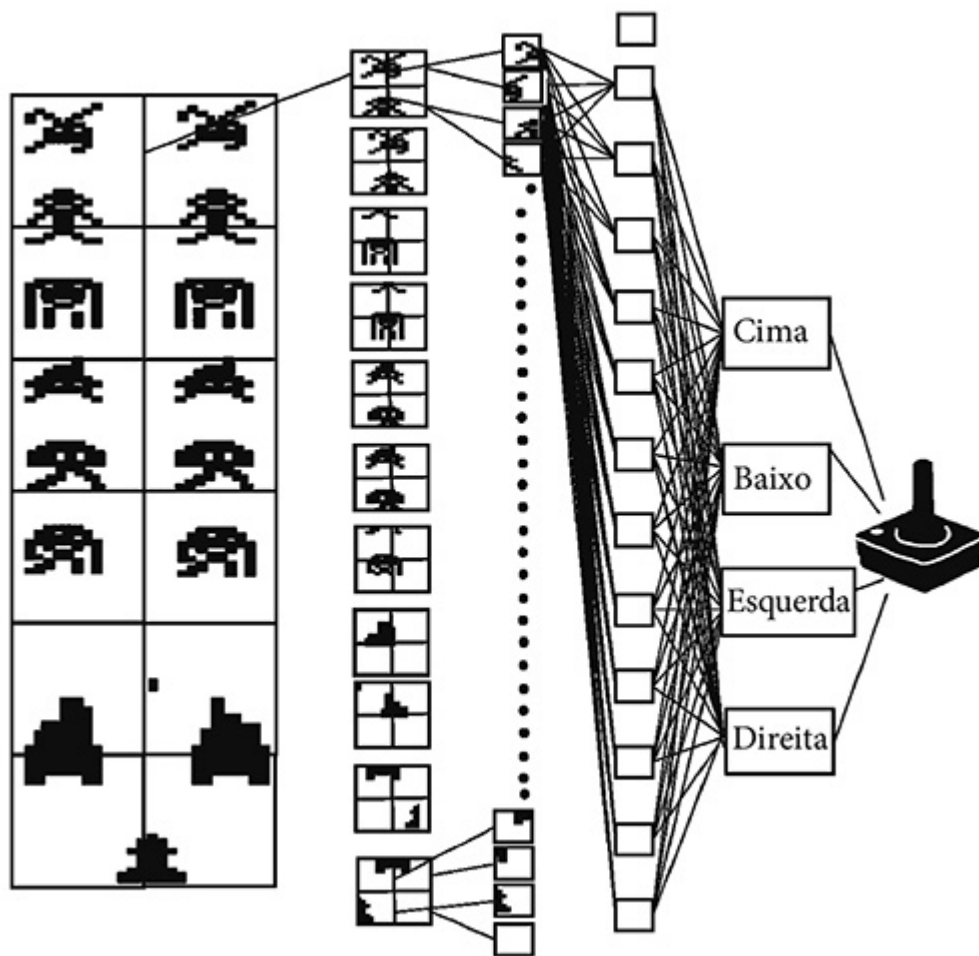


Figura 16.1 Funcionamento interno de uma rede neural convolucional. Figura feita por Elise Sumpter.

As camadas mais profundas da rede neural colocam as pequenas imagens juntas novamente, alimentando-as por meio de uma quarta e quinta camadas de neurônios completamente conectados (Figura 16.1). É dentro dessas camadas que a rede aprende como relacionar as imagens

pequenas e a melhor ação a se realizar. Também é aqui que ela calcula o tamanho dos padrões importantes do jogo. Se a rede estiver aprendendo sobre um jogo em que os objetos importantes são grandes, como os boxeadores do *Boxing*, então a rede irá formar várias conexões muito parecidas entre os neurônios que estiverem próximos uns aos outros. Se os objetos forem pequenos, como os invasores espaciais ou os hexágonos que o *Q*bert* deseja pintar, então as conexões serão mais complexas. Se os objetos puderam mudar de tamanho, como os tanques quando ficam mais próximos no *Robotank*, então as conexões serão parecidas entre camadas convolucionais diferentes. Redes neurais convolucionais são uma técnica poderosa porque encontram o tamanho e a forma de padrões importantes automaticamente, sem que os programadores tenham que dizer o que procurar.

A ideia de redes neurais convolutas existe desde os anos 1990, mas por um longo tempo elas foram apenas uma, dentre várias diferentes, competindo com algoritmos que foram propostos para ajudar computadores a detectarem padrões. Foi em 2012, quando Alex Krizhevsky fez uma apresentação de quatro minutos na conferência de Sistemas de Processamento de Informação Neural no Lago Tahoe, na Califórnia, que os pesquisadores começaram a dar uma atenção especial ao método convolucional. Alex havia entrado numa competição, que acontece todo ano como parte da conferência, para identificar automaticamente diferentes objetos em imagens.³ Para um humano, essa é uma tarefa mundana. As imagens são uma variedade de cenas: alguns homens exibindo o peixe grande que eles pescaram; uma fila de carros esportivos; um bar cheio de gente e duas mulheres fazendo uma selfie são apenas algumas de uma variedade de cenas descrevendo muitos aspectos diferentes da vida. A tarefa é encontrar os objetos — o peixe, os homens, os carros, as pessoas no bar e o telefone usado para fazer a selfie. Antes de Alex apresentar seu trabalho na conferência, essa tarefa de identificação de imagem tinha se mostrado muito difícil. Até mesmo os algoritmos que haviam sido ajustados ao problema cometeram erros pelo menos uma entre quatro vezes.

Alex reduziu essa taxa de erro para uma em seis usando uma rede neural convolucional. Ele não falou muito para o algoritmo sobre o tamanho e a forma dos objetos que deveria classificar. Apenas deixou o algoritmo aprender. E quando foi abastecido com milhões de imagens, ele aprendeu muito bem mesmo. Enquanto outras abordagens dependiam de humanos para determinar as características importantes — como bordas, formas e contrastes de cor — que definem objetos, a abordagem de Alex era simplesmente configurar a rede e deixá-la fazer o trabalho.

A competição de 2012 foi só o começo. Alex era um estudante de doutorado usando um par de placas gráficas de jogos instaladas em seu computador. Uma vez que a técnica foi revelada, outros pesquisadores a refinaram e usaram computadores mais poderosos para o trabalho. No ano seguinte, o participante ganhador cometeu erros em apenas um entre oito casos e em 2017 a taxa de erro ficou abaixo de 2%.⁴ Tanto o Google quanto o Facebook começaram a notar esses fatos. Eles perceberam que redes neurais convolucionais solucionavam problemas no centro dos seus negócios. Um algoritmo que consegue reconhecer automaticamente os rostos de nossos amigos, nossos animais fofos preferidos e os lugares exóticos que já visitamos pode permitir a essas empresas atingir com mais eficiência os nossos interesses.

Alex e seu orientador de doutorado, Geoffrey Hinton, foram contratados pelo Google. No ano seguinte, um dos vencedores da competição, Rob Fergus, recebeu uma oferta de emprego do Facebook. Em 2014, o Google montou seu próprio time ganhador e recrutou rapidamente a doutoranda de Oxford Karen Simonyan, que ficou em segundo lugar. Em 2015, foi o pesquisador da Microsoft Kaiming He e seus colegas que ganharam o prêmio. Kaiming foi recrutado pelo Facebook no ano seguinte.⁵ Um a um, os principais pesquisadores de redes neurais foram levados para trabalhar nos grupos de inteligência artificial recém-formados na Microsoft, no Google e no Facebook.

Esses pesquisadores não foram apenas recrutados para encontrar objetos em imagens. O avanço importante da pesquisa de Alex foi ter demonstrado

que redes neurais convolucionais poderiam aprender a resolver um problema sem terem que ser “informadas” sobre qual problema estariam resolvendo. Logo ficou claro que a mesma abordagem derrotaria seus competidores no reconhecimento de caligrafia e de pronúncia. Ela poderia ser utilizada para reconhecer ações em filmes curtos e prever o que aconteceria a seguir.

Por essa razão, as redes neurais convolucionais que jogaram os jogos do Atari eram muito mais empolgantes do que o algoritmo que havia derrotado Kasparov no xadrez. Quando o Deep Blue venceu, os pesquisadores haviam estabelecido que um computador poderia derrotar um humano em um jogo difícil. Quando a competição terminou e a mídia havia encerrado suas entrevistas, a máquina foi desligada e os pesquisadores voltaram para trabalho do dia a dia.

A chegada das redes neurais foi diferente. Os algoritmos solucionavam um problema atrás do outro. O reconhecimento de faces no iPhone X da Apple utiliza as redes neurais para identificar unicamente o rosto de seu dono. A Tesla utiliza as redes neurais no sistema de visão de seus carros para avisar sobre possíveis colisões.⁶ As conquistas das redes neurais convolucionais em problemas de visão foram acompanhadas de aprimoramentos parecidos das redes recorrentes em problemas de linguagem. O Google realizou grandes progressos na qualidade de suas traduções do inglês para o chinês por meio do uso de novas técnicas de rede neural.⁷

A escala potencial de aperfeiçoamentos em algoritmos baseados em redes neurais não está clara. No mínimo, pesquisas recentes aumentaram em muito a habilidade de computadores em reconhecer objetos, sons e frases. Mas o entusiasmo em torno desse método sugere que estamos a caminho de resultados ainda mais dramáticos. Será que estamos finalmente perto de solucionar o problema de criar máquinas inteligentes de um modo geral?

Eu certamente gostaria de responder a essa pergunta geral de IA. Mas antes de eu poder chegar perto de solucionar uma pergunta tão grande, senti que precisava responder a uma pergunta menor antes. Quão boas são as

redes neurais convolucionais? O artigo sobre a análise de imagem de Alex Krizhevsky havia revolucionado tanto a indústria quanto a academia, mas eu queria entender melhor os limites das redes neurais.

Apesar de toda a propaganda, uma boa parte da resposta já pode ser encontrada na própria pesquisa do Google. Voltemos à rede neural que joga Atari. Os pesquisadores descobriram que ela consegue jogar 29 dos 49 jogos testados no mesmo nível de humanos, derrotando os profissionais na maioria deles. Isso significa que os humanos são melhores do que as redes neurais em 20 dos jogos testados. Em alguns deles, o desempenho da rede neural foi apenas um pouco melhor do que a aleatoriedade.

Quando olhei a lista de desempenhos computador *versus* humano, um jogo em particular chamou minha atenção: *Ms Pac-Man*. O desempenho da rede neural nesse jogo de labirinto simples, no qual a *Ms Pac-Man* tem que mastigar a maior quantidade de pastilhas possível ao mesmo tempo que evita ser pega por fantasmas, foi muito ruim. Ela conseguiu pontuar cerca de 12% dos pontos de um jogador profissional.

O jogo *Ms Pac-Man* é mais complicado do que *Breakout* ou *Space Invaders* porque ele envolve paciência. Se a *Ms Pac-Man* quiser sobreviver, deverá ser cautelosa, esperar que uma área não contenha fantasmas e então se deslocar para lá. As “pastilhas poderosas”, que tornam os fantasmas azuis e permitem a *Ms Pac-Man* comê-los, devem ser usadas com moderação. Se ela as tomar muito cedo no jogo, ganhará alguns pontos ao comer fantasmas, mas terá dificuldades mais tarde, quando os fantasmas voltarem ao normal e a caçarem, quatro contra um.

As redes neurais convolucionais simplesmente não conseguem lidar com esses aspectos do jogo. Elas só conseguem reagir ao que elas enxergam diretamente à sua frente: alienígenas que precisam ser abatidos, um boxeador que precisa ser socado e hexágonos que precisam ser pisados. Elas não conseguem planejar o futuro, mesmo que em curto prazo. O algoritmo fracassa em todos os jogos do Atari que requerem um planejamento mínimo do que vem a seguir.

A estrutura dessas redes significa que elas são boas em identificar objetos em imagens, juntar sons para formar palavras e reconhecer qual o próximo passo em um jogo de tiros. Mas não podemos esperar que elas façam algo a mais do que essas tarefas, nem elas. A inteligência artificial mais avançada da atualidade consegue ver coisas e produzir respostas imediatas, mas não consegue entender o que está olhando. Ela não consegue bolar um plano.

Não fui a única pessoa cuja atenção foi capturada pela *Ms Pac-Man*. O pesquisador da Microsoft Harm van Seijen me disse que, depois de ler o artigo do Google, o jogo chamou sua atenção. Ele ficou surpreso de a rede neural não ter conseguido solucionar o problema e se perguntou por que diferia do *Space Invaders*.

Ele e seus colegas desenvolveram uma abordagem alternativa. Determinaram que o problema de *Ms Pac-Man* seria resolvido mais facilmente quando decomposto em problemas menores. Eles modelaram todos os componentes do jogo — as pastilhas, as frutas e os fantasmas — como agentes que competiam pela atenção da *Ms Pac-Man*. Depois treinaram a rede neural para equilibrar a força com a qual esses agentes a “atraíam” em pontos diferentes do labirinto. O resultado foi uma *Ms Pac-Man* muito cautelosa, que levou um tempo maior para completar os níveis, mas nunca foi comida. O algoritmo deles acumulou uma pontuação de 999.900, e, nesse momento, o jogo travou.

Os pesquisadores que desenvolvem redes neurais estão perfeitamente cientes do problema de “dizer muito para seus algoritmos”. O objetivo a longo prazo do programa de inteligência artificial geral é treinar redes com o mínimo de entradas humanas possível. Se queremos que os algoritmos exibam as qualidades que vemos em inteligência humana ou animal, então essas redes devem aprender sozinhas, sem termos que lhes dizer o que fazer ou a quais padrões prestar atenção.

Por outro lado, esses mesmos pesquisadores estão interessados em demonstrar que podem ganhar em jogos de computador mais complexos e mais modernos. O desafio máximo é o *StarCraft*, um videogame de estratégia sofisticada jogado competitivamente como um e-sport. O time do

DeepMind da Google configurou, juntamente com a empresa criadora do jogo, a Blizzard, um ambiente de software para ajudar pesquisadores a construir um algoritmo para aprender a jogar *StarCraft II*.⁸ Esse ambiente fornece uma representação abstrata do jogo, em vez da visão de tela mostrada a jogadores humanos, possibilitando aos programadores pular o desafio de reconhecer os objetos na tela.

Harm foi muito cético sobre a possibilidade de aprender jogos de computador modernos a partir da visão pixelada de tela sem fornecer ao algoritmo informações adicionais. Ele me disse: “Aprender *StarCraft* do zero? Isso não vai acontecer.” Ele disse que aprender *StarCraft* irá exigir que os programadores forneçam informações específicas do jogo e que usem uma rede neural altamente especializada.

A própria abordagem de Harm a *Ms Pac-Man* se situa entre a abordagem de redes neurais pura para os jogos de Atari “mais fáceis” e o trabalho em *StarCraft*. Diferentemente da rede neural que joga *Space Invaders*, Harm forneceu, à sua rede neural, a posição da Ms Pac-Man, dos fantasmas e das pastilhas. Mas ele não informou à rede, de antemão, que as pastilhas eram boas ou que os fantasmas eram maus. Em vez disso, a rede aprendeu como equilibrar o risco de se deslocar em direção a um fantasma, com os benefícios de se mover em direção à comida. Ela então aprendeu como juntar essas informações.

Quando falei com Harm, ele insistia no desafio de equilibrar o quanto de informação ele deveria dar à sua rede neural. “Estou interessado em como fazer com que computadores aprendam comportamentos por meio da interação com o mundo”, ele me disse. “Eu não especifico o que um agente [como a Ms Pac-Man] precisa fazer, eu especifico unicamente o que ela quer realizar.” Para Harm, esse é o desafio fundamental para as pesquisas, em vez de simplesmente conseguir que um computador obtenha uma maior pontuação em *Ms Pac-Man*.

O desafio de descobrir o quanto um computador consegue aprender do zero é central ao entendimento do quão longe estamos de criar uma IA geral. Após ter falado com Harm, contatei, em outubro de 2017, David

Silver, que lidera a equipe do DeepMind que está treinando redes neurais para jogar *Go*.

O algoritmo AlphaGo que a equipe de David criou havia derrotado o jogador de *Go* número um do mundo, Ke Jie, em maio de 2017. O algoritmo havia iniciado sua vida aprendendo um manual de jogadas contendo 30 milhões de movimentos feitos pelos maiores jogadores de *Go* do mundo. Ele então aperfeiçoou suas habilidades jogando contra várias variações de si próprio. O algoritmo final combinava uma enciclopédia de conhecimento sobre o jogo, uma rede neural que havia aprendido jogando e uma máquina de calcular poderosa que procurava possíveis movimentos. É um feito impressionante de engenharia, mas é um algoritmo altamente especializado.

David também havia se envolvido com o projeto sobre jogos de Atari, então achei que ele teria conhecimento sobre o equilíbrio entre aprender do zero e construir um algoritmo especializado, como havia feito com o *Go*. Eu lhe enviei, por e-mail, uma série de perguntas, mas ele me respondeu pedindo que eu tivesse paciência, pois um “artigo novo em poucas semanas” responderia minhas perguntas.

A espera valeu a pena. Em 19 de outubro de 2017, David e sua equipe publicaram um artigo na revista *Nature* descrevendo o AlphaGo Zero, um novo algoritmo para jogar *Go* que derrotava todos os algoritmos anteriores. Não apenas isso, esse algoritmo funcionava sem a assistência humana. Eles configuraram uma rede neural, deixaram que jogasse vários jogos de *Go* contra si mesma, e alguns dias depois ela era a melhor jogadora de *Go* do mundo.

Fiquei impressionado, muito mais do que quando um computador venceu no xadrez ou no pôquer, ou até mesmo mais do que com o primeiro campeão de *Go* do David. Dessa vez, não havia manuais nem algoritmo de busca especializado, apenas uma máquina jogando partida após partida, de iniciante a nível de mestre, até um nível que ninguém jamais, computador ou pessoa, conseguiria derrotar. Era uma rede neural aprendendo a jogar, por si própria, um jogo muito complexo.

David me disse que não acreditava haver qualquer motivo para que sua abordagem não fosse também capaz de aprender a jogar *Ms Pac-Man*, apesar de o Google ainda não ter tentado. Para David, o novo campeão de *Go* respondia muitas das dúvidas sobre quão amplamente as redes neurais do DeepMind conseguiam aprender para resolver diferentes problemas do zero. Ele e seus colegas enfatizaram que, “dadas as regras do jogo, [a rede neural] aprende por tentativa e erro”.

Quando eu indaguei Harm sobre o novo campeão de *Go*, ele me disse que isso era “definitivamente impressionante e um claro avanço”, mas não estava convencido de que isso era aprender do zero.

Harm me disse que “as regras dos jogos de Atari são desconhecidas pelo algoritmo e precisam ser aprendidas”. O desafio em aprender a jogar jogos de computadores de baixo para cima é entender como a tela muda em resposta a uma ação. Essa informação já é fornecida ao computador para o *Go*.

Tomas Mikolov, o engenheiro do Facebook que desenvolveu redes neurais de processamento de linguagem, não via o *Go* como um passo em direção a uma IA mais geral. “Há um perigo de superotimização de algo irrelevante para o objetivo final [de criar uma IA] caso se trabalhe com problemas altamente artificiais como jogos de computadores ou *Go*, xadrez etc.”, ele me disse. Tomas acredita que é somente por intermédio do trabalho com tarefas baseadas em linguagem, com as quais nos comunicamos e ensinamos um sistema de IA, que conseguiremos desenvolver inteligência verdadeira.

Eu me lembrei da minha própria infância. Quando vi *Ms Pac-Man* ou *Space Invaders* pela primeira vez, não precisei de mais do que alguns minutos para entender o que estava acontecendo. Não me parecia estranho ou não natural ser capaz de controlar as formas na TV, e meu cérebro rapidamente compreendeu as conexões entre o joystick e a tela. Rapidamente eu estava alcançando altas pontuações de um cartucho para outro. Testemunhei a mesma coisa quando meu filho de 12 anos abriu o *Overwatch* da Blizzard pela primeira vez no Playstation 4. Sua mente

interpretou imediatamente as formas em movimento, compreendeu como os controles funcionavam e começou a raciocinar sobre os elementos estratégicos do jogo.

Quando as crianças aprendem a jogar jogos de computador, usam uma estratégia bem diferente daquela adotada no interior de uma rede neural. Elas chegam ao jogo com um modelo já estabelecido de como objetos interagem, tanto na vida real quanto dentro de jogos. Brenden Lake, da Universidade de Nova York, em colaboração com colegas do MIT e de Harvard, examinou cuidadosamente um jogo do Atari chamado *Frostbite*. Eles perceberam que conseguiam aprender a jogar o jogo muito mais rapidamente que uma rede neural, já que assimilaram bem rápido os objetivos do jogo e como os objetos se movem. Eles precisaram de apenas dois minutos observando um vídeo no YouTube de alguém jogando e mais 15 a 20 minutos de tempo de jogo para elevar sua pontuação ao mesmo nível de uma rede neural.

Brenden propôs, ainda, alguns experimentos mentais interessantes. Ele percebeu que conseguia facilmente adaptar seu estilo de jogar para, por exemplo, conseguir o máximo de peixes possível no jogo ou continuar o máximo de tempo sem morrer. Para a rede neural, completar essas tarefas implicaria começar seu treinamento novamente do zero. Apesar de os cientistas entenderem bastante sobre o cérebro, continuamos espetacularmente incapazes de simular o entendimento espontâneo de novos contextos dos humanos.

Google, Tesla, Amazon, Facebook e Microsoft estão todos em uma corrida para produzir esse entendimento. Muito de seu trabalho é colaborativo; eles compartilham bibliotecas de código e compartilham os últimos conhecimentos em conferências, como a Nips [sigla em inglês para Sistemas de Processamento de Informação Neural]. Há uma verdadeira agitação sobre o progresso realizado e uma rivalidade amigável entre os grupos. Alguns desses engenheiros e matemáticos acreditam que estejamos cerca de apenas uma década da verdadeira IA. Outros a veem a séculos de distância.

Eu fico imaginando o que os chefes deles acham de todo este
excitamento.

O CÉREBRO BACTERIANO

Em 2016, o CEO do Facebook, Mark Zuckerberg, propôs a si mesmo um desafio: construir um mordomo automatizado que o ajudaria com as tarefas domésticas. O nome de seu mordomo, Jarvis, foi inspirado por um robô de inteligência artificial construído pelo personagem Homem de Ferro dos quadrinhos e da série de filmes *Os vingadores*. O Jarvis da ficção tem uma inteligência como a dos humanos, capaz de ler os pensamentos do Homem de Ferro e se conectar emocionalmente com ele. Jarvis combina um imenso banco de dados de informações com o poder de raciocinar e assimilar conceitos. Mark não precisava de um assistente que o ajudasse a salvar o mundo, mas quis verificar quão inteligente um assistente doméstico conseguiria criar usando a biblioteca de algoritmos do Facebook.

O Google tem ambições que vão além de mordomos pessoais. O time do DeepMind, que ganhou no *Go* e no *Space Invaders*, tem ajudado o Google a melhorar a eficiência de energia em seus servidores e a desenvolver um discurso mais realístico para o assistente pessoal da empresa. Outra área de aplicação é a DeepMind Health, um projeto sobre o qual um dos googlers de Londres havia me falado quando eu os visitei. O objetivo é examinar como o Serviço Nacional de Saúde do Reino Unido coleta e administra os dados dos pacientes a fim de verificar como o processo pode ser melhorado. O CEO da DeepMind, Demis Hassabis, fala que, um dia, sua equipe vai produzir um

“artigo científico de alta qualidade em que o primeiro autor será uma IA”. O objetivo final é substituir muitos dos difíceis desafios intelectuais executados por engenheiros, doutores e cientistas por soluções criadas por máquina inteligentes.

Os pesquisadores que trabalham nesses projetos com frequência afirmam que estão progredindo lentamente em direção a uma inteligência artificial mais geral. Muitos deles acreditam que estão nos levando numa viagem cada vez mais próxima à chamada “singularidade”, o ponto no qual os computadores se tornam tão inteligentes quanto nós. A hipótese da singularidade afirma que, uma vez que esse ponto seja alcançado, uma vez que os computadores sejam capazes de desenvolver outras máquinas inteligentes e aperfeiçoarem a si próprios, então nossa sociedade mudará dramaticamente e para sempre. As máquinas podem até nos considerar supérfluos.

Em um encontro, em janeiro de 2017, do Instituto do Futuro da Humanidade — uma organização filantrópica em Boston, Massachusetts, que se preocupa em lidar com riscos futuros —, o físico teórico Max Tegmark organizou uma comissão para debater a inteligência artificial geral.¹ A comissão incluiu nove dos homens mais influentes no campo, incluindo o empreendedor e CEO da Tesla, Elon Musk; o guru do Google, Ray Kurzweil; o fundador da DeepMind, Demis Hassabis; e Nick Bostrom, o filósofo que mapeou nosso caminho para o que ele chama de “superinteligência”.

As opiniões dos membros da comissão variavam sobre se uma inteligência de máquina em nível humano aconteceria gradual ou repentinamente, ou se será boa ou ruim para a humanidade. Mas todos concordaram que uma forma geral de IA seria mais ou menos inevitável. Eles também acharam que isso está suficientemente próximo de acontecer, ao ponto de termos que começar a pensar, já, sobre como iremos lidar com isso.

Apesar da convicção da comissão de que a IA está a caminho, meu ceticismo aumentou enquanto eu observava a fala deles. Eu havia passado o

ano anterior dissecando algoritmos utilizados dentro das empresas que esses caras dirigem e, pelo que pude observar, eu simplesmente não consegui entender de onde eles acham que essa inteligência virá. O que encontrei nos algoritmos que eles estão desenvolvendo é insuficiente para sugerir que uma inteligência como a humana esteja a caminho.

Pelo que pude ver, essa comissão, consistindo em um “quem é quem” da indústria tecnológica, não estava levando a questão a sério. Eles estavam curtindo a especulação, mas não era ciência. Era puro entretenimento.

O problema com minha opinião — falar sobre o que acontecerá quando a IA geral chegar é pura especulação — é que ela é difícil de se provar. Parte de mim sente que eu nem deveria tentar. Se eu tentar, vou simplesmente me juntar ao clamor de homens de meia-idade desesperados para ter suas opiniões ouvidas. Mas parece que não consigo me controlar. Com Stephen Hawking afirmando que a IA “poderia pressagiar o fim da raça humana”, não consigo evitar querer esclarecer minha própria opinião.

Tentei argumentar contra a probabilidade de uma IA geral antes. Em 2013, participei de um debate on-line com Olle Häggström, professor da Universidade de Gotemburgo, sobre o assunto.² Olle acredita que o risco é grande o suficiente para que devêssemos nos certificar de que a humanidade esteja preparada para sua chegada. Precisamos garantir que minimizemos os riscos de qualquer transição em potencial para a superinteligência.

Meu contra-argumento para a preparação excessiva atual para uma IA é que ela é apenas um dentre todos os tipos de riscos desconhecidos do futuro. Ainda estamos tendo dificuldades com o aquecimento global. Vivemos numa época em que uma guerra nuclear poderia começar com poucas horas de aviso. Se pensarmos daqui a cem anos, existe uma possibilidade real de que nosso planeta seja atingido por um grande meteoro ou uma enorme erupção solar, que um vulcão escureça os céus por muitos anos e/ou que entremos numa era do gelo severa. Esses são desafios que temos conhecimento e para os quais devemos nos preparar.

Existem muitas outras ameaças que podem surgir com a tecnologia. Suponha que biólogos, acidentalmente (ou não), desenvolvam

geneticamente um supervírus ou uma espécie de bactéria mortal que infecte e lentamente mate mamíferos, nós inclusive.

Imagine o que aconteceria se encontrássemos um modo de prolongar a vida humana indefinidamente de forma que ninguém precisasse morrer. A demanda por recursos seria imensa, e os conflitos, inevitáveis. Pense nos riscos que poderiam ser criados por causa do surgimento de uma gosma cinzenta de nanopartículas. Essa gosma poderia ser criada por cientistas numa escala bem pequena, e desenvolver a habilidade de se reproduzir e “comer” tudo no nosso planeta.

Se formos tão ficção científicos quanto a IA geral, então precisamos também considerar como deveríamos nos preparar para a descoberta de inteligência alienígena quando o Telescópio Espacial James Webb começar a tirar fotos mais detalhadas de nosso universo. O que faremos se descobrirmos que as estrelas estão se movendo de um modo que contradiz as regras da física e só puderem ser explicadas por inteligência extraterrestre? O que pensar sobre a teoria de *Matrix* que afirma que estamos todos vivendo numa simulação de computador? Não deveríamos investir mais em pesquisas que verifiquem anomalias em potencial em nossa realidade?

Se qualquer uma dessas coisas acontecer antes da IA geral, então elas representarão um risco tão grande para a humanidade quanto um computador superinteligente. Ironicamente, muitos desses cenários apocalípticos não são de minha autoria, mas foram documentados cuidadosamente por Olle. Eu os descobri por intermédio de nossas discussões e da leitura de seu excelente livro *Here Be Dragons*.³

Olle, no entanto, não se convenceu com meu argumento. Ele ainda considera que a IA geral seja um dos grandes riscos para a continuação da nossa existência ou, no mínimo, um risco que deveria ser controlado com cuidado.

Decidi tentar uma abordagem diferente dessa vez. Uma abordagem menos filosófica e mais prática. Vou me concentrar em esclarecer onde

estamos agora. Daí, vou deixar vocês (e o Olle) tirarem suas próprias conclusões sobre o que o futuro nos reserva.

Em outra discussão na mesma conferência do debate de Max Tegmark, Yan Le Cun, inventor das redes neurais convolucionais e pesquisador principal de IA do Facebook, descreveu que resolver o problema da identificação de imagens, usando seu método, foi como escalar uma montanha.⁴ Agora eles haviam superado este desafio e estavam de volta a um vale olhando para o próximo pico. Yan não sabia quantas montanhas a mais havia para serem escaladas, mas achou que poderia haver “outras 50”. Demis Hassabis, da DeepMind, estimou o número de montanhas como sendo menor que 20. Em sua opinião, cada uma dessas montanhas compreende uma lista de problemas não resolvidos sobre como estimulamos diferentes propriedades conhecidas do cérebro.⁵

A metáfora da montanha levanta mais perguntas do que responde. Como eles podem saber, só de olhar para a próxima montanha, que ela pode ser conquistada? Eles não têm um mapa do território nem uma rota clara que os leve a cada cume. E, pior ainda, como sabem que não vão escalar uma dessas montanhas e encontrar um pico que simplesmente não pode ser alcançado? O que escalar uma montanha nos diz sobre a próxima?

Uma boa maneira de colocar o desenvolvimento algorítmico em um contexto é pensar sobre as tarefas que esses algoritmos conseguem e as que não conseguem fazer, e por quê. Durante a sessão de perguntas e respostas que sucedeu a discussão no Instituto do Futuro da Humanidade, Anca Dragan, pesquisadora assistente de engenharia elétrica e ciência da computação de Berkeley, que era, de um modo geral, mais cética sobre os desenvolvimentos da IA do que vários outros participantes, deu um exemplo de um desafio aparentemente simples que parecia pouco provável de ser solucionado num futuro próximo. Ela disse que não ficaria surpresa se robôs fossem incapazes de esvaziar uma lavadora de louças empilhada por humanos em alguns anos futuros.

Outro cético, Oren Etzioni, CEO do Instituto Allen de Inteligência Artificial, usou a linguagem como exemplo. Ele disse à plateia que

computadores “não conseguem determinar com segurança ao que palavras como ‘isso’ se conectam numa frase. E *isso* é bastante frustrante para as pessoas que acreditam que [os computadores] estão prestes a dominar o mundo”.

Sabemos que o “isso” em sua segunda frase se refere à implicação geral da primeira frase. Até mesmo os melhores algoritmos de linguagem são incapazes de determinar se o “isso” se refere a uma frase, um computador, à própria palavra “isso” ou a alguma outra ideia dentro do contexto da frase.

O meu exemplo preferido do que a IA não consegue fazer vem do futebol. Melhores momentos de jogos recentes podem ser encontrados online. Dois robôs se movendo repetidamente um em direção ao outro enquanto a bola fica parada a meio metro de distância de ambos; um goleiro robô que não consegue fazer uma defesa enquanto uma bola vinda de um chute fraco passa lentamente por ele; jogadores caindo repetidamente por causa da força de seus próprios chutes, que os impulsiona para baixo. Assistir a esses robôs ilustra o quanto ainda precisamos melhorar.

Depois da Copa do Mundo de Robôs de 2016, entrevistei Tim Laue e Katie Genter, que eram, respectivamente, membros do time campeão de Bremen, na Alemanha, e do time vice-campeão de Austin, no Texas, na classe robótica clássica. Eles me disseram que se concentram em escrever algoritmos que detectam as linhas do campo, as traves dos gols, a bola e os outros jogadores. Cada algoritmo usado é para resolver um problema específico: detectar a bola ou o movimento do chute. Essa abordagem de cima para baixo está bem distante da IA de baixo para cima que seria essencialmente requerida se os robôs quisessem derrotar oponentes humanos. Os robôs estão realizando uma sequência de tarefas de identificação, em vez de aprender como jogar o jogo. Ainda falta muito até que tenhamos um jogador de futebol da IA.

Como a IA ainda não consegue realizar tarefas humanas, será que ela consegue competir com animais? Podemos diminuir ligeiramente nossas expectativas e criar um algoritmo que seja tão esperto quanto um cachorro?

Algumas pessoas, incluindo muitos cientistas que deveriam estar mais bem informados, falam sobre animais em termos de reações simples do tipo estímulo-resposta. O exemplo clássico é o cachorro de Pavlov salivando ao som de um sino. Qualquer um que tenha um cachorro lhe dirá que essa visão pavloviana é uma grande simplificação, e eles estão certos. A visão típica que um dono de cachorro tem sobre seus bichos de estimação serem como membros da família ou amigos não é apenas uma visão emocional e antropocêntrica. Está de acordo com a visão da maioria dos biólogos behavioristas modernos sobre animais domésticos — como compartilhando muitos de nossos comportamentos complexos. Juliane Kaminski, chefe do projeto de cognição canina da Universidade de Southampton, descobriu que cachorros conseguem aprender de um modo parecido com o que as crianças pequenas aprendem, levando em consideração as perspectivas do mundo de seus donos para decidir quais objetos pegar, e entender nossas intenções a partir dos nossos movimentos corporais.⁶

Essas qualidades, de entender o contexto de situações diferentes e aprender sobre como aprender, continuam sendo perguntas em aberto na pesquisa de IA. Até que tenhamos feito progressos muito mais significativos em direção ao modelamento humano do que fizemos até agora, não seremos capazes de simular cachorros, gatos ou qualquer outro animal doméstico.

Talvez eu tenha almejado muito alto com cães, então vamos descer alguns níveis para os insetos e as abelhas em particular. Lars Chittka, da Universidade Queen Mary de Londres, documentou recentemente nossa compreensão crescente sobre a cognição das abelhas e descobriu que abelhas possuem um intelecto fantástico.⁷ Após uns poucos voos circulando seus ninhos, abelhas recém-nascidas adquirem uma boa ideia de como seu mundo se parece. Elas, então, rapidamente começam a trabalhar coletando comida. Abelhas trabalhadoras aprendem o cheiro e a cor das melhores flores e resolvem o “problema do caixeiro viajante”, o de visitar as fontes de comida disponíveis no menor tempo possível. Elas conseguem se lembrar onde se sentiram ameaçadas e, algumas vezes, “veem fantasmas” ao reagirem a uma impressão de perigo que não existe na verdade. Abelhas que

encontram muita comida se tornam otimistas e começam a subestimar o perigo de ataques predadores. A rede neural intrínseca, modelada pelo cérebro da abelha, tem uma estrutura bem diferente das redes neurais artificiais convolucionais ou recorrentes. As abelhas parecem ser capazes de reconhecer a diferença entre objetos usando apenas quatro neurônios de entrada e aparentam não possuir qualquer representação interna de imagens. Por outro lado, outras tarefas mais simples do tipo estímulo-resposta, que poderiam ser modeladas em termos de uma quantidade pequena de portas lógicas, afetam regiões inteiras do cérebro.

A coisa mais extraordinária sobre as abelhas é que elas conseguem aprender a jogar futebol!!! Bem, não exatamente futebol, mas um jogo muito parecido.⁸ O grupo de pesquisa de Lars treinou abelhas para empurrarem uma bola através de um gol. As abelhas conseguiram aprender a realizar essa tarefa de várias maneiras diferentes, incluindo observar uma abelha de plástico empurrar a bola e observar outras abelhas verdadeiras completarem a tarefa. Elas não precisaram praticar o jogo extensivamente a fim de aprender a tarefa. Rolamento de bola não é algo com que abelhas se deparam normalmente em suas vidas, então o estudo mostra que as abelhas conseguem aprender novos comportamentos rapidamente, sem a necessidade da abordagem repetitiva de tentativa e erro. É exatamente esse o problema que as redes neurais artificiais não conseguiram resolver até agora. As abelhas conseguem generalizar suas habilidades em outras áreas para vencer um novo problema, como o futebol.

É importante lembrar que o problema da inteligência artificial geral não é sobre se os computadores são ou não melhores do que os humanos em certas tarefas. Já vimos que um computador consegue jogar jogos como xadrez, Go e pôquer melhor que os humanos. Então não acho que as máquinas teriam muitos problemas para derrotar uma abelha nesses jogos. A pergunta é se podemos, em um computador, produzir aprendizado de baixo para cima do tipo largamente observado em animais. No momento, as abelhas são capazes de generalizar seus entendimentos do mundo de um modo que os computadores não conseguem.

A minhoca nematoide *C. elegans* é um dos animais vivos mais simples. Um adulto completamente desenvolvido consiste em 959 células, das quais cerca de 300 são neurônios. Isso comparado com as 37.200.000.000.000 células em seu corpo⁹ e os 86.000.000.000 neurônios em seu cérebro.¹⁰ As *C. elegans* são largamente estudadas porque, apesar de sua simplicidade relativa, compartilham muitas de nossas propriedades, incluindo comportamento, interações sociais e aprendizagem.

Monika Scholz, da Universidade de Chicago, recentemente criou um modelo de como a minhoca usa inferência probabilística, parecida com aquela usada no modelo de pesquisa eleitoral de Nate Silver, no capítulo 8, para decidir quando se mover.¹¹ A minhoca “pesquisa” seu meio ambiente local para medir quanta comida há disponível e, então, “prevê” se é melhor ficar parada ou começar a explorar para obter novos recursos. Estudos como esse revelam detalhes da tomada de decisões pela minhoca, mas eles ainda não modelam o organismo por inteiro. Um projeto separado, conhecido por OpenWorm, tenta capturar aspectos do mecanismo de como as minhocas se movem, mas mais trabalho é necessário para unir esses modelos e reproduzir o repertório completo do comportamento da *C. elegans*. No momento, não sabemos realmente como as 959 células agem juntas e, então, não podemos modelar propriamente o comportamento de um dos animais mais simples da Terra.

Então vamos esquecer, pelo menos por enquanto, isso de criar a inteligência de cachorros, abelhas, minhocas e, até mesmo, jogadores de futebol. Que tal uma ameba? Podemos reproduzir a inteligência de microrganismos?

O bolor limoso, *Physarum polycephalum*, é um organismo ameboide que constrói redes minúsculas de tubos para transportar nutrientes entre diferentes partes de seu corpo. Audrey Dussutour, da Universidade de Toulouse na França, mostrou que os bolores limosos se habituariam à cafeína, uma substância que eles normalmente tentam evitar, e depois voltam a seus comportamentos normais quando lhes é dada a opção de evitar a substância.¹² Outros estudos mostraram que os bolores limosos conseguem

antecipar eventos periódicos, escolhem uma dieta balanceada, navegam ao redor de armadilhas e constroem redes que conectam diferentes fontes de comida eficientemente. O limo pode ser interpretado como uma forma de computador distribuído, recebendo sinais de partes diferentes de seu corpo e tomando decisões baseadas em experiências anteriores. Ele faz tudo isso sem um cérebro e sem um sistema nervoso.

Pode até ser possível produzir um modelo matemático compreensível de bolores limosos num futuro próximo, mas certamente ainda não chegamos lá. A “memória” e o aprendizado do limo poderiam ser potencialmente modelados por um tipo de componente elétrico conhecido por “memristor”, uma combinação entre um resistor e capacitor que fornece uma forma de memória flexível.¹³ Mas nós ainda não sabemos como configurar uma rede de memristores para combiná-los uns com outros e solucionar problemas de um modo que imite o bolor limoso.

A etapa imediatamente abaixo do bolor limoso na complexidade biológica são as bactérias. A bactéria *E. Coli* é um “bichinho” que vive em nossas entranhas. Apesar de a maioria das espécies ser benigna ou até mesmo benéfica, algumas delas nos dão intoxicação alimentar. A *E. coli* e outras bactérias navegam por nosso corpo, absorvem os açúcares e “decidem” como crescer e quando se dividir.¹⁴ Elas são altamente adaptáveis. Quando você toma um copo de leite, os genes para a absorção da lactose são ativados dentro da *E. coli*, mas se você comer em seguida uma barra de chocolate, os genes que processam a glucose, “preferida” pela *E. coli*, suprimem os genes da lactose. A bactéria move-se através de um movimento do tipo correr e rolar, correndo em uma direção antes de rolar para “escolher” uma nova direção. Elas sintonizam essas taxas de “rolamento” com a qualidade do meio ambiente em que se encontram. Cada um dos diferentes “objetivos” da bactéria — obtenção de recursos, locomoção e reprodução — estão equilibrados por meio de diferentes combinações de genes ativados e desativados.

O equilíbrio, na *E. coli*, dos diferentes objetivos na busca pela obtenção de recursos não parece familiar? Deveria parecer. A Ms Pac-Man é uma

bactéria. As tarefas que esses dois organismos, real e artificial, almejam completar são muito parecidas. A fim de se adaptarem, ambas têm que responder aos sinais de entrada de várias fontes diferentes: a *E. coli* regula o consumo de recursos, responde a perigos e navega por obstáculos. Os neurônios da Ms Pac-Man respondem aos fantasmas, às pastilhas de comida e à estrutura do labirinto. Os corpos nos quais as bactérias vivem não são idênticos, assim como os labirintos da Ms Pac-Man diferem uns dos outros, mas os algoritmos que cada uma delas emprega são flexíveis o bastante para lidar com uma gama ampla de desafios ambientais.

Eu havia encontrado o equivalente biológico mais próximo ao nível mais elevado da IA atual. É um bichinho de barriga.

Um argumento contra minha analogia bactéria-cérebro é que a razão pela qual não conseguimos simular minhocas e bolores limosos é que não sabemos o que esses organismos desejam alcançar. Alguns dos pesquisadores de redes neurais com os quais conversei argumentaram que nós não sabemos o que esses especialistas chamam de função objetivo das minhocas. Para treinar uma rede neural, precisamos ser capazes de informar-lhe qual padrão ela se destina a produzir e, em teoria, se sabemos o padrão, i.e., a função objetivo, deveríamos ser capazes de reproduzi-lo. Existe alguma validade nesse argumento — biólogos não têm um entendimento completo da *C. elegans* nem dos bolores limosos.

Em última análise, entretanto, o argumento “nos diga a função objetivo” desvia do problema verdadeiro. O trabalho experimental de biólogos sobre inteligência revela mais sobre *como* o cérebro trabalha, entendendo as conexões entre neurônios e os papéis das diferentes partes do cérebro, do que revela sobre o padrão global de *por que* nosso cérebro tem objetivos específicos. Se os neurocientistas vão trabalhar juntamente com especialistas da inteligência artificial para criar máquinas inteligentes, então esse trabalho conjunto não pode depender de os biólogos encontrarem a função objetivo de animais e informá-la para os especialistas de aprendizado de máquina. Progressos em IA devem envolver biólogos e cientistas da computação trabalhando juntos para entender os detalhes do cérebro.

Os testes de IA deveriam, na minha opinião, se basear naquele primeiro que foi proposto por Alan Turing, em seu famoso teste “jogo da imitação”.¹⁵ Um computador passa no teste de Turing, ou jogo da imitação, se conseguir convencer um humano, durante uma sessão de perguntas e respostas, de que ele é, de fato, um humano. Esse é um teste difícil e estamos longe de alcançar esse objetivo, mas podemos usar o teste principal de Turing como ponto de partida para uma série de testes mais simples.

Numa seção menos bem citada de seu artigo de 1950, Turing propõe simular uma criança como um passo em direção a simular um adulto. Poderíamos nos considerar “aprovados” em um miniteste de jogo da imitação quando estivermos convencidos de que o computador é uma criança. Meu argumento é que deveríamos utilizar a rica diversidade de organismos do nosso planeta como uma série de casos de testes.¹⁶ Podemos reproduzir a inteligência exibida pelo bolor limoso, por minhocas e abelhas num modelo de computador? Se conseguirmos capturar seus comportamentos quando eles estão se movimentando nos ambientes deles e interagindo uns com os outros, então poderemos afirmar que produzimos um modelo das inteligências gerais deles. Até conseguirmos produzir esses modelos, deveremos ser cautelosos com as afirmações que fazemos. Baseado em evidências atuais, estamos modelando inteligência num nível parecido ao de uma única bactéria.

Bem... não exatamente. Harm van Seijen foi bastante cauteloso ao explicar que seu algoritmo de *Ms Pac-Man* não poderia ser considerado como tendo sido construído do zero. Ele o ajudou ao dizer-lhe para prestar atenção aos fantasmas e às pastilhas. Em contrapartida, o conhecimento da bactéria sobre os perigos e recompensas em seu ambiente tem sido construído de baixo para cima, por meio da evolução.

Harm me disse: “Muitas das pessoas que estão falando sobre IA estão otimistas demais; elas subestimam o quão difícil é construir os sistemas.” Foi baseado em sua experiência ao desenvolver *Ms Pac-Man* e outros sistemas de aprendizado de máquina que ele percebeu que estamos realmente bem longe de uma forma geral de IA.

Mesmo que consigamos criar uma inteligência totalmente bacteriana, Harm foi cético sobre quanto mais longe podemos chegar. Ele disse: “Os humanos são realmente bons em reutilizar o que foi aprendido ao realizar uma determinada tarefa em outra tarefa parecida; nossos atuais algoritmos de ponta são terríveis nisso.”

Tanto Harm da Microsoft quanto Tomas Mikolov do Facebook viram um risco em dar às redes neurais nomes sofisticados e fazer grandes afirmações.

O fundador da empresa em que Harm trabalha atualmente parece concordar com ele. Em setembro de 2017, Bill Gates disse ao *Wall Street Journal* que o tema IA não é algo que precisamos temer. Ele disse que discordava de Elon Musk sobre a urgência dos problemas em potencial.

Então, se estamos, no momento, imitando um nível de “inteligência” parecido com o de um bichinho de barriga, por que Elon Musk declarou que a IA é uma grande preocupação? Por que Stephen Hawking estava ficando tão preocupado com o poder preditivo de seu software de fala? O que faz Max Tegmark e seus colegas enfileirarem-se para declarar, um após o outro, que a superinteligência está a caminho? Essas pessoas são inteligentes; o que está nublando seus julgamentos?

Eu acho que existe uma combinação de fatores. Um é comercial. Não atrapalha em nada a DeepMind ter um pouco de burburinho sobre inteligência artificial. Demis Hassabis suavizou a ênfase em “resolução de inteligência” que sua empresa tinha quando o Google adquiriu o DeepMind, e em entrevistas recentes manteve o foco na solução de problemas matemáticos de otimização. O trabalho sobre Go demonstra que a DeepMind tem a vanguarda em problemas como descoberta de drogas e otimização nas redes de energia, que requerem computação pesada para encontrar a melhor solução dentre várias alternativas disponíveis. Se não fosse essa propaganda publicitária desde cedo, a DeepMind poderia não ter adquirido os recursos para solucionar alguns desses problemas importantes.

Elon Musk não suavizou sua retórica. Ele parece estar adotando uma posição de continuar exagerando deliberadamente a IA para impulsionar

muitos de seus projetos superambiciosos, incluindo veículos autônomos. Esses projetos de longo prazo somente podem obter sucesso se os clientes estiverem dispostos a comprar a ideia de que adquirir o mais recente carro Tesla é um passo em direção a um futuro fantástico. Frequentemente é um sonho distante que impulsiona nosso desejo de descobrir coisas sobre o mundo.

Não é o dinheiro em si que motiva as pessoas da DeepMind ou Tesla. Existe também um sentimento genuíno de excitação entre os pesquisadores. Na virada do milênio, a ideia da IA geral parecia estar morta. Quando redes neurais solucionaram o problema de classificação de imagem em 2012, pareceu que uma mudança estava finalmente por vir.

A verdade por trás dos algoritmos atuais é frequentemente muito mais simples e mais banal do que o termo “inteligência artificial” pode dar a entender. Quando analisei os algoritmos que tentam nos classificar, descobri que eram representações estatísticas de mais ou menos as mesmas coisas que já sabemos sobre nós mesmos. Quando analisei os algoritmos que tentaram nos influenciar, descobri que eles estavam explorando alguns aspectos muito simples de nosso comportamento para decidir quais informações de busca nos mostrar e o que tentar nos vender. Redes neurais decifraram alguns jogos, mas não conseguimos enxergar ainda um caminho até o topo da próxima montanha. Quando Alex e eu construímos nosso próprio bot de linguagem, ele conseguiu nos enganar com algumas frases antes de se revelar um fracasso total.

Após um ano de desafios para construir Jarvis, o mordomo, Mark Zuckerberg colocou um vídeo no Facebook mostrando o resultado. Vou lhe avisando antes de você assistir a ele que é bajulação pura. Somos apresentados ao quarto de Mark quando ele acorda em sua tradicional camiseta cinza. Jarvis informa a Mark sobre suas reuniões do dia e lhe comunica que já estava divertindo sua filha de um ano com “lições de mandarim” desde que ela acordou. Quando Mark desce até a cozinha, a torradeira já está ligada, tendo antecipado que chegou o momento da sua fatia diária. Depois há um toque de campainha no sistema da porta da

frente, causado pelo software de reconhecimento facial ao identificar os rostos de seus pais à medida que eles se aproximam da porta. O dia termina com ele e sua esposa interagindo com um sistema de som que toca uma música ambiente.

Zuckerberg foi bem direto numa publicação em um blog sobre o que Jarvis consegue e não consegue fazer.¹⁷ Na verdade, seu desafio em programar um mordomo reflete muitos dos desafios que eu mesmo enfrentei ao dissecar algoritmos usados na sociedade. Seu produto final é uma aula magna em aplicações práticas das técnicas que vimos neste livro: reconhecimento de voz e de face por meio de redes neurais convolucionais, um tipo de modelo de regressão usado para prever quando ele deseja uma torrada ou quando sua filha deseja aprender mandarim e algoritmos “também gostaram” para escolher a música que toda a família deseja escutar. O Facebook desenvolveu uma biblioteca de rotinas de programação que ajudam seus empregados, e, neste caso, Mark Zuckerberg, a transformar esses algoritmos em aplicações.

Quando superei o sentimentalismo do vídeo e investiguei profundamente seu blog, percebi que Zuckerberg é bastante esperto (sei que deve ser um pouco tarde para dizer isso). Ele chega ao resultado como um alquimista de dados. Enquanto seus colegas do Silicon Valley ficam tentando provar suas qualificações intelectuais sentando-se em comissões de discussão recepcionadas por físicos teóricos, Mark está confinado programando interfaces e descobrindo como tirar o maior proveito possível das ferramentas que sua empresa criou. A conclusão a que ele chegou faz sentido: “A IA está mais próxima de ser capaz de fazer coisas bastante poderosas do que a maioria das pessoas supõe — dirigir carros, curar doenças, descobrir planetas, compreender meios de comunicação. Cada uma dessas coisas terá um grande impacto no mundo, mas ainda estamos decifrando o que é inteligência real.”

Há possibilidades fantásticas de desenvolver produtos e serviços utilizando os algoritmos que analisamos neste livro. Esses algoritmos irão continuar a transformar nosso lar, nosso local de trabalho e a forma pela

qual viajamos, mas ainda estão bem distantes de uma IA geral. As técnicas estão dando um tipo de inteligência bacteriana a nossas torradeiras, nossos sistemas de som caseiros, nossos escritórios e nossos carros. Esses algoritmos têm o potencial de reduzir as tarefas domésticas que precisamos fazer, mas eles não se parecerão nada com humanos.

Kathleen Richardson, pesquisadora de ética e cultura de robôs e IA, da Universidade de Montfort em Leicester, chama os progressos recentes obtidos em algoritmos de uma “inteligência publicitária” em vez de “inteligência artificial”. Ao falar com a BBC World Service, ela disse que “o que realmente aconteceu nos últimos dez anos foi que as grandes corporações se aperfeiçoaram em coletar dados de consumidores e vender produtos de volta a esses consumidores”. O mordomo de Mark Zuckerberg é um exemplo perfeito disso. Mark coleta um carregamento de dados sobre o que escuta no Spotify, os rostos de seus amigos e sua rotina diária, alimenta isso tudo em um computador e o ajuda a melhorar sua vida cotidiana.

O verdadeiro perigo, segundo Kathleen, não é que a inteligência de computador exploda, mas, em vez disso, que estamos utilizando as ferramentas que temos atualmente para melhorar a vida de uns poucos, em vez da vida de muitos. Há o perigo de nos concentrarmos em mordomos para os super-ricos, em vez de soluções para todos os outros.

Max Tegmark e seus amigos têm o direito de se reunirem e especularem sobre o futuro da IA, mas sua discussão deveria ser vista pelo que ela é. Um bando de homens ricos, com mais ou menos a mesma formação socioeconômica, educação e experiência de trabalho, que querem conversar entre si sobre ficção científica. Quando discuti esses assuntos com Olle Häggström, caí na mesma armadilha.

De volta ao mundo real, os humanos ainda continuarão sendo a única forma de inteligência humana por um longo tempo. A grande pergunta é: iremos usar os algoritmos que já temos para ajudar uma sociedade mais abrangente ou servir às necessidades de poucos? Eu sei qual dessas duas opções eu preferiria.

DE VOLTA À REALIDADE

Estou em pé, com um pequeno grupo numa festa.

Estou apenas esperando que aconteça.

A mesma coisa que me aconteceu em quase toda reunião social no ano passado. Um período durante o qual passei a maior parte do tempo em meu escritório: programando algoritmos, ajustando modelos estatísticos, escrevendo sobre meus resultados e entrevistando engenheiros e cientistas por Skype.

A conversa é fútil. Ouço educadamente. Então alguém menciona. Não é a mesma coisa toda vez. As histórias são ligeiramente diferentes, mas o tema é o mesmo.

“O perigo com o Facebook é que ele controla o que as pessoas veem”, ele diz. “Li sobre um estudo que fizeram em que mostravam somente publicações negativas e as pessoas ficaram deprimidas. Ele consegue controlar nossas emoções, e essas vibrações negativas se espalham como um vírus.”

Eu permaneço quieto e deixo outra pessoa falar. “Sim. Eles lucram com a venda dos dados dos usuários. É bem capaz de o Trump ter vencido as eleições nos Estados Unidos porque sua empresa de Oxford ou algum outro lugar baixou os perfis de todos”, ela diz.

“Veja bem, o problema é que os algoritmos deles foram treinados com notícias falsas”, diz o próximo. “O mesmo acontece com o Google. Eles construíram um computador para compreender linguagem e o computador começou a dizer que odiava muçulmanos.”

“Eu sei”, e o primeiro homem volta à conversa, “e isso é apenas o início. Os cientistas calculam que daqui a vinte anos haverá um computador com inteligência humana que irá decidir nos transformar em cliques de papel. Eu li isso no livro do Elon Musk.”

Foi então que explodi. “Primeiramente”, eu disse, “o estudo do Facebook mostrou que pessoas bombardeadas com notícias negativas eram propensas a escrever uma palavra negativa a mais por mês, um efeito minúsculo, mas estatisticamente significativo. Em segundo lugar, a empresa, chamada Cambridge Analytica, não teve nada a ver com o sucesso eleitoral do Trump. O CEO deles havia feito uma série de afirmações sem verificação sobre o que eles eram capazes de fazer. Em terceiro lugar, sim, computadores treinados em nossa linguagem fazem, de fato, associações sexistas e racistas, mas isso porque nossa sociedade é implicitamente preconceituosa. A maioria das buscas do Google dá a vocês, e ao restante da população, os resultados mais desinteressantes e acurados imagináveis. O maior problema com esse serviço é que ele está sobrecarregado com links sem sentido tentando fazê-lo ir à Amazon e comprar mais porcarias. E, finalmente, houve alguns progressos recentes em redes neurais, mas a pergunta sobre se conseguiremos criar uma IA geral permanece em aberto. Não conseguimos sequer fazer um computador aprender a jogar *Ms Pac-Man* direito.”

Todos olharam para mim. “Ah, e o Elon Musk é um idiota”, emendei.

Eu me odeio. Eu odeio o babaca chato em que me transformei. Este idiota entediante que leu todos os artigos científicos, que tem que estragar tudo com detalhes e advertências. Eu nem tenho tanta coisa assim contra o Elon Musk. Ele simplesmente está fazendo o trabalho dele. Adicionei isso para que a conversa pudesse voltar a uma módica despreocupação.

Sei que julguei mal a situação. Estou sendo pedante e mesquinho. Além das pequenas imprecisões nos detalhes, as pessoas com as quais eu estava

falando têm preocupações genuínas sobre como a sociedade está mudando. É por isso que eles estavam falando sobre o Facebook, Cambridge Analytica e inteligência artificial.

Imediatamente após minha visita ao Google, faz um ano agora, eu havia sentido o mesmo que eles. Eu havia sentido medo do futuro que os matemáticos e os cientistas da computação estavam criando para nós.

Eu compreendo agora que estes algoritmos não são assim tão assustadores, do jeito que eu havia imaginado que fossem. É uma pena que algoritmos não tenham resolvido os problemas que a nossa sociedade tem com sexismo e racismo, mas eles tampouco pioraram as coisas. Eles sinalizaram problemas sobre tendências que todos nós precisamos nos esforçar mais para resolver. É assustador que o grupo SCL possa criar uma empresa como a Cambridge Analytica, que se vende como se mirasse em personalidades, quando ela não possui as ferramentas ou os dados para tal, mas isso aí é o capitalismo global. Aceite-o ou certifique-se de atacar o sistema e não suas mentiras e trapanças. É deprimente que o Facebook contenha tantas notícias falsas e o Twitter esteja cheio de bots trolls, mas é reconfortante saber que quase ninguém está dando ouvidos a eles. É preocupante que o sucesso on-line seja apenas parcialmente relacionado a talento, mas é um consolo quando falhamos. É um pouco decepcionante que você precise ser bonito para se dar bem no Tinder, mas isso não é diferente do que já acontece, certo?

Compreender os algoritmos nos permite entender os cenários futuros um pouco melhor. Se você consegue entender como os algoritmos funcionam hoje, então será mais fácil julgar quais cenários são realistas e quais não são. O grande risco que encontrei com os algoritmos é quando não pensamos racionalmente sobre seus efeitos, quando nos deixamos levar por cenários de ficção científica.

Estou consciente agora de que minha jornada, aqui em pé após minha tentativa fracassada de responder ao que era apenas uma brincadeirinha de “experimente o Facebook” e uma especulação de inteligência artificial “o que

isso tudo significa”, me transformou numa pessoa não muito interessante para se ter uma conversa.

Felizmente, sou casado. Minha mulher começou a falar de *Pokémon Go*. Lovisa às vezes desaparece para encontrar homens de 25 anos de idade no parque para caçar Pokémon. E agora ela conta para o grupo como eles se reuniram na semana passada para tentar capturar o Moltres: minha mulher, os homens jovens em calças de moletom, pais de seus 30 anos com bebês em carrinhos, um grupo de garotos que por acaso passava por ali e um casal de idosos que passa por esse caminho todos os dias como parte de sua caçada diária ao Pokémon. Ela conseguiu o Moltres. A maioria deles conseguiu, mas uma das crianças menores começou a chorar quando o Pokémon pulou para fora da bola e voou para longe. O casal de idosos confortava o menino. Eles irão tentar novamente amanhã.

“Você deveria calcular como a probabilidade de capturar o Moltres muda com o número de pokebolas que eu arremesso”, sugere Lovisa. As outras pessoas no grupo em que estou agora concordam com a cabeça. Uma concordância um pouco exagerada, percebo. Por que o David esteve desperdiçando seu tempo com pedantismo algorítmico quando poderia estar nos iluminando com a teoria de jogos do *Pokémon Go*?

Eu não acho que vá me importar com *Pokémon*. Melhor deixar algumas coisas sem análise. Ninguém quer saber a probabilidade de uma criança começar a chorar.¹

Em vez disso, penso sobre aquelas pessoas no parque mostrando, umas às outras, suas cartas de *Pokémon* e os espaços em branco indicando os que ainda faltam caçar. Pessoas que jamais iriam se encontrar para se divertir juntas.

Penso na minha filha de 14 anos e meio, Elise. Ela saiu semanas atrás para encontrar um amigo de um grupo no Snapchat. Lovisa e eu ficamos preocupados no começo — seria esse amigo on-line um pedófilo de 40 anos de idade? Nossas preocupações não se justificaram. Elise se encontrou com um menino normal de 13 anos de idade com cabelo azul brilhante. Este verão ela quer ir visitar outro amigo que conheceu on-line e que vive na

Polônia. Eles se falam frequentemente pelo Skype enquanto fazem seus deveres de casa juntos. Os pais de Elise terão que pensar bastante sobre se permitirão isso ou não.

Meu filho, Henry, e eu acabamos de retornar de uma viagem a Newcastle. Pelo Twitter eu havia entrado em contato com outro pai, Ryan, que, assim como eu, treina o time de futebol do seu filho. Eu havia organizado uma viagem para levar 32 meninos de 12 anos de idade da Suécia para jogar uma série de jogos contra a equipe dele.

Ryan também me convidou para falar sobre o *Soccermatics* em seu trabalho, o Departamento de Trabalho e Pensões (DWP). O lugar — um austero edifício ministerial nos arredores da cidade — era bem diferente do da minha palestra no Google em Londres no ano anterior. Mas dentro, eu me vi cercado de cientistas de dados em uma sala pequena com o mesmo entusiasmo para visualizar e compreender dados.

Após minha palestra, Ryan me mostrou o que ele e sua equipe (isto é, a do trabalho) estavam fazendo. Eles querem conectar conselhos locais no Reino Unido que enfrentam dificuldades parecidas, como um grande empregador local fechando as portas, de modo que consigam tirar proveito de experiências compartilhadas. A abordagem de Ryan utilizava estatística de ponta, mas era focada em ajudar seus colegas da DWP que poderiam, em troca, ajudar as pessoas nas comunidades locais. “Queremos possibilitar que as autoridades se familiarizem com os dados”, ele me disse, “para que descubram as soluções por eles mesmos.”

A visão de Ryan era de que precisamos combinar as inteligências humana e algorítmica para resolver nossos problemas. Algoritmos sozinhos não eram suficientes.

Eu me lembrei de algo que Joanna Bryson havia me dito. Eu havia perguntado a ela sobre a ascensão dos algoritmos e se ela achava que eles iriam se tornar tão espertos quanto nós. Ela me disse que eu estava fazendo a pergunta errada. “Já ultrapassamos a inteligência humana de muitas formas”, ela disse.

Nossa cultura havia produzido formas de inteligência “artificial” matemática para resolver problemas ao longo de milhares de anos: desde a geometria dos babilônios e dos egípcios, passando pelo desenvolvimento do cálculo por Newton e Leibniz, a calculadora de mão, que nos permitiu fazer uma aritmética mais rápida, até o computador moderno, a sociedade conectada e o nosso mundo algorítmico da atualidade. Estamos ficando mais e mais espertos com a ajuda dos modelos matemáticos, e os modelos estão se aprimorando porque nós os desenvolvemos. Os algoritmos são parte de nossa bagagem cultural. Somos parte deles e eles são partes de nós.

Este patrimônio está sendo construído ao nosso redor, nos lugares mais improváveis, incluindo, parece, o Departamento de Trabalho e Pensões do Reino Unido.

Há riscos com a forma por meio da qual interagimos on-line. Muita coisa não é bonita, mas há tantas possibilidades incríveis para acontecer ao se trabalhar juntamente com algoritmos. E, por ora, pelo menos, somos nós que controlamos os algoritmos e não eles que nos controlam. Estamos dando forma à matemática a nossa própria imagem.

Domingo pela manhã, um grupo de jovens de Gateshead jogou futebol contra um grupo de jovens de Uppsala. Era um dia ensolarado, com um bom trabalho de equipe e fair play e pais felizes torcendo pelos seus meninos. Depois, todos assistimos juntos ao espetáculo de fogos de artifício no clube de rúgbi local. Tudo porque um algoritmo havia sugerido a Ryan e a mim que seguíssemos um ao outro pelo Twitter.

Percebi que eu estava sonhando acordado e voltei à conversa. O assunto ainda era o Facebook. Então eu disse: “Você sabia que as recomendações de anúncio do Facebook incluem categorias para pessoas interessadas em torrada e ornitorrinco?”

“Não no mesmo sanduíche, eu espero”, disse a mulher ao meu lado.

Todo mundo ri.

Já não me sinto mais tão dominado pelos números.

Notas

Capítulo 1: Procurando Banksy

1. www.fortune.com/2014/08/14/google-goes-darpa
2. www.dailymail.co.uk/femail/article-1034538/Graffiti-artist-Banksy-unmasked-public-schoolboy-middle-class-suburbia.html
3. www.bbc.com/news/science-environment-35645371

Capítulo 2: Gerando ruído

1. O'Neil, C. 2016. *Weapons of Math Destruction: How big data increases inequality and threatens democracy* [Armas de destruição matemática: Como o Big Data aumenta a desigualdade e ameaça a democracia]. Crown Books.
2. Você pode verificar a configuração de sua conta do Google em: www.google.com/settings/u/0/ads/authenticated
3. www.thinkwithgoogle.com/intl/en-aunz/advertising-channels/video/campbells-soup-uses-googles-vogon-to-reach-hungry-australians-on-youtube/
4. Você pode baixar o plug-in no site www.noiszy.com
5. *Weapons of Math Destruction* exemplifica esses tipos de discriminação em detalhes.
6. Datta, A., Tschantz, M. C. e Datta, A. 2015. “Automated experiments on ad privacy settings” [Experimentos automatizados sobre configurações de privacidade em anúncios]. *Proceedings on Privacy Enhancing Technologies* 2015 [Anais sobre tecnologias de aprimoramento da privacidade], nº 1: 92–112.

7. www.post-gazette.com/business/career-workplace/2015/07/08/Carnegie-Mellon-researchers-see-disparity-intargeted-on-line-job-ads/stories/201507080107
8. Amit e seus colegas discutem aqui possíveis explicações: www.fairlyaccountable.org/adfisher/#disc-cause
9. www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
10. www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race
11. Veja www.ajlunited.org
12. O trabalho de Jonathan Albright foi parte de um artigo do *Guardian* sobre preenchimento automático e notícias falsas: www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook
13. Burrell, J. 2016. “How the machine ‘thinks’: Understanding opacity in machine learning algorithms” [“Como a máquina ‘pensa’: entendendo a opacidade dos algoritmos de aprendizado de máquina”]. *Big Data & Society* [Big Data & Sociedade] 3, nº 1: 2053951715622512.

Capítulo 3: Os principais elementos da amizade

1. Reuni as últimas 15 publicações das páginas dos meus amigos do Facebook em 13 de dezembro de 2016. A maioria das publicações pôde ser categorizada, e aquelas que não pudeam deixei de fora da análise.
2. Começamos construindo o gráfico da primeira categoria em função de cada uma das outras 12 categorias. Depois, construímos o gráfico da segunda categoria em função das outras 11 restantes (i.e., excluindo a primeira). A terceira ficou em função de dez categorias e assim por diante. No total, foram $12 + 11 + 10 + \dots + 1 = (13 \times 12)/2 = 78$ gráficos ao todo.
3. Descrevi a análise de componentes principais no texto principal de uma perspectiva geométrica. Para aqueles de vocês que estudaram álgebra, observo que o ACP é encontrado calculando-se primeiro a matriz covariante das 13 categorias. A primeira componente principal é, então, o autovetor (que é uma reta) associado ao maior autovalor da matriz covariante. A segunda componente principal é o autovetor associado ao segundo maior autovalor e assim por diante.

Capítulo 4: Suas 100 dimensões

1. Kosinski, M., Stillwell, D. e Graepel, T. 2013. “Private traits and attributes are predictable from digital records of human behavior” [Traços pessoais e atributos são conjecturados a partir de registros digitais do comportamento humano]. *Proceedings of the National Academy of Sciences* [Anais da Academia Nacional de Ciências dos Estados Unidos] 110, nº 15: 5802–5.
2. Kosinski, M., Wang, Y., Lakkaraju, H. e Leskovec, J. 2016. “Mining big data to extract patterns and predict real-life outcomes” [Explorando o Big Data para extrair padrões e prever resultados da vida real]. *Psychological methods* [Métodos psicológicos] 21, nº 4: 493.
3. Costa, P. T. e McCrae, R. R. 1992. “Four ways five factors are basic” [As quatro maneiras em que cinco fatores são básicos]. *Personality and individual differences* [Personalidade e diferenças individuais] 13, nº 6: 653–65.
4. <https://research.fb.com/fast-randomized-svd>
5. Os pesquisadores do Facebook descrevem um método que eles implementaram no seguinte artigo: Szlam, A., Kluger, Y. e Tygert, M. 2014. “An implementation of a randomized algorithm for principal component analysis” [Uma implementação de um algoritmo aleatório para a análise de componentes principais]. *arXiv preprint arXiv:1412.3510*. Os detalhes de como eles usam esses algoritmos na prática não estão, em geral, disponíveis.
6. www.npr.org/sections/alltechconsidered/2016/08/28/491504844/you-think-you-know-me-facebook-but-you-dont-know-anything
7. Loius, J. J. e Adams, P. “Social dating” [Encontro amoroso]. US Patent 9.609.072, emitida em 28 de março de 2017.
8. Kluemper, D. H., Rosen, P. A. e Mossholder, K. W. 2012. “Social networking websites, personality ratings, and the organizational context: More than meets the eye?” [Sites de relacionamento pessoal, avaliações de personalidade e o contexto organizacional: Além do que se pode ver?]. *Journal of Applied Social Psychology* [Revista de Psicologia Social Aplicada] 42, nº 5: 1143–72.
9. Stéphane, L. E. e Seguela, J. “Social networking job matching technology” [A tecnologia da oferta e procura de empregos em redes sociais]. US Patent Application 3/543.616, solicitada em 6 de julho de 2012.
10. Veja as seguintes patentes:
 - Donohue, A. “Augmenting text messages with emotion information” [Ampliando mensagens de texto com informações sentimentais]. US Patent Application 14/950,986, solicitada em 24 de novembro de 2015.

- MATAS, Michael J., Reckhow, M. W., e Taigman, Y. “Systems and methods for dynamically generating emojis based on image analysis of facial features” [Sistemas e métodos para gerar emojis dinamicamente baseado na análise de expressões faciais em imagens]. US Patent Application 14/942.784, solicitada em 16 de novembro de 2015.
- Yu, Y., e Wang, M. “Presenting additional content items to a social networking system user based on receiving an indication of boredom” [Apresentar informações adicionais ao Sistema de relacionamento pessoal do usuário baseado no recebimento de uma indicação de tédio]. US Patent 9.553.939, emitida em 24 janeiro de 2017.

11. Hibbeln, M. T., Jenkins, J. L., Schneider, C., Joseph, V. e Weinmann, M. 2016. “Inferring negative motion from mouse cursor movements” [Deduzindo postura negativa a partir dos movimentos do mouse].

12. Hehman, E., Stoler, R. M. e Freeman, J. B. 2015. “Advanced mouse-tracking analytic techniques for enhancing psychological science” [Técnicas analíticas avançadas de rastreamento de mouse para melhorar a ciência da psicologia]. *Group Processes & Intergroup Relations* [Processamento de Grupos & Relações de Intergrupo] 18, nº 3: 384–401.

Capítulo 5: A hiperbolitização de Cambridge

1. O artigo mais recente foi o tema de um recurso judicial da companhia: www.theguardian.com/technology/2017/may/14/robert-mercator-cambridge-analytica-leave-eu-referendum-brexit-campaigns

2. Este é, claro, um exemplo de tal afirmação binária. Eles são difíceis de se evitar.

3.

https://d25d2506sfb94s.cloudfront.net/cumulus_uploads/document/0q7lmn19of/TimesResults_160613_EUReferendum_W_Headline.pdf

4. A pesquisa de opinião não tem a idade exata das pessoas entrevistadas, então, ao adaptar o modelo, supus que cada pessoa tinha a média das idades declaradas. Usei regressão logística com uma função de ligação logit, com os pesos proporcionais aos tamanhos das populações. Para testar a solidez, também analisei as funções quadráticas de idade, mas isso não melhorou o ajuste do modelo significativamente.

5.

https://d25d2506sfb94s.cloudfront.net/cumulus_uploads/document/0q7lmn19of/TimesResults_160613_EUReferendum_W_Headline.pdf

6. Skrandal, A. e Rabe-Hesketh, S. 2003. “Multilevel logistic regression for polytomous data and rankings” [Regressão logística multinível para dados politômicos e classificações]. *Psychometrika* [Psicometria] 68, nº 2: 267–87.
7. www.ca-political.com/services/#services_audience_segmentation
8. www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data
9. www.youtube.com/watch?v=n8Dd5aVXLCC
10. Aqui só incluo usuários com mais de 50 “curtidas” em categorias que tiveram um total de mais de 150 “curtidas”.
11. www.theintercept.com/2017/03/30/facebook-failed-to-protect-30-million-users-from-having-their-data-harvested-by-trump-campaign-affiliate
12. Uma série de denúncias feitas por Carole Cadwalladr, do *Guardian*, e por repórteres do *Channel 4*, de notícias, revelou uma gama de supostas práticas duvidosas de gerenciamento de dados na companhia. Mas quando os efeitos colaterais foram controlados, ainda não havia evidências de que Nix e seus colegas tinham criado um algoritmo de predição de personalidade. Enquanto o delator Christopher Wylie alegou que ele e Alex Kogan tinham ajudado a CA a construir uma ferramenta de “conflito psicológico”, os detalhes da eficácia desta ferramenta propriamente dita não foram revelados. A falta de uma prova concreta confirmou minha própria análise e, com a avaliação de Alex Kogan — os dados do Facebook ainda não são suficientemente detalhados para permitir uma análise apropriada que permita a construção de propagandas direcionadas às inclinações individuais das pessoas. Com o anúncio da queda das ações do Facebook em 7% por causa de uma mera possibilidade de uma ferramenta de “conflito psicológico”, acho que o gigante da internet vai ser bem cuidadoso quanto a controlar o acesso de terceiros aos dados dos usuários do Facebook no futuro.

Capítulo 6: Impossivelmente imparcial

1. Vários outros artigos acusaram o Compas de ser uma caixa-preta, mas o artigo — Brennan, T., Dieterich, W. e Oliver, W. 2004. “The Compas scales: Normative data for males and females. Community and incarcerated samples” [As escalas do Compas: dados normativos para homens e mulheres. Amostras da sociedade e dos detidos]. Traverse City, MI: Northpointe Institute for Public Management — fornece uma descrição bem completa de como o modelo é construído a partir de regressão e decomposição em valor singular. A página 147 desse documento fornece a equação chave usada para prever reincidência.

2. Brennan, T., Dieterich, W. e Ehret, B. 2009. Evaluating the predictive validity of the Compas risk and needs assessment system” [Avaliando a validade previsível do risco do Compas]. *Criminal Justice and Behavior* [Justiça criminal e comportamento], 36, nº 1: 21–40.
3. www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
4. Dieterich, W., Mendoza, C. e Brennan, T. 2016. Compas risk scales: Demonstrating accuracy equity e predictive parity [Escalas de risco do Compas: Demonstrando acurácia na equidade e paridade preditiva]. Technical report, Northpointe, July 2016. www.northpointeinc.com/northpointe-analysis.
5. www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
6. Corbett-Davies, S., Pierson, E., Feller, A., Goel, S. e Huq, A. (2017) “Algorithmic decision making and the cost of fairness.” [A decisão algorítmica e o preço da justiça]. Em Proceedings of the 23^o ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Anais da 23^a ACM SIGKDD Conferência Internacional sobre Descoberta do Conhecimento e Exploração de Dados], p. 797–806. ACM.
7. O anúncio de oferta de emprego varia de país para país. O recurso “publicar oferta de emprego” só é implementado nos Estados Unidos e no Canadá, e os anúncios são verificados para discriminações em potencial. Então, na prática, talvez não seja possível que o anúncio que sugeri seja implementado pelos monitores do Facebook, e você deve tratar o texto nessa seção como um experimento idealizado. Contudo, realmente encontrei uma publicação “relevante” de um membro da equipe de ajuda do Facebook de como direcionar por intermédio de demografia, idade e gênero usando um método muito menos sutil do que o que descrevo agora: www.facebook.com/business/help/community/question/?id=10209987521898034
8. Na versão completa de minha explicação pedante, também observo que a proporção de mulheres que não viram o anúncio, mas que estavam interessadas no emprego, i.e., $50/550 = 9,09\%$, é igual à proporção de homens que estavam interessados no emprego, mas que não viram o anúncio, i.e., $50/550 = 9,09\%$. Os números da tabela foram escolhidos deliberadamente para criar um resultado em que a propaganda é igualmente confiável para ambos os grupos.
9. Kleinberg, J., Mullainathan, S. e Raghavan, M. 2016. “Inherent trade-offs in the fair determination of risk scores” [Compensações inerentes na determinação de avaliações de risco]. *arXiv preprint arXiv:1609.05807*.
10. A idade média de acusados brancos no banco de dados do Condado de Broward é 35, enquanto a idade média de acusados negros é 30.
11. Identifiquei essa relação usando um modelo de regressão logística para prever dois anos de taxas de reincidência para o banco de dados do Condado de Broward coletado pela ProPublica. A variável

dependente foi se o acusado reincidiu no período de dois anos. Dividi os dados em um conjunto de dados de treinamento (90% de observações) e um conjunto de dados de testes (10% de observações). Usando o conjunto de testes, descobri que o modelo logístico que melhor previu a reincidência era baseado na idade ($\beta_{\text{idade}} = -0,047$; P-valor $< 2e-16$) e o número de antecedentes ($\beta_{\text{antecedentes}} = 0,172$; P-valor $< 2e-16$), combinado com uma constante ($\beta_{\text{constante}} = 0,885$; P-valor $< 2e-16$). Isso significa que acusados mais velhos são menos propensos a serem presos por crimes futuros, enquanto aqueles com um número maior de antecedentes são mais propensos a serem presos novamente. A raça não foi, estatisticamente, um prognosticador significativo de reincidência (num modelo multivariado incluindo raça, um fator afro-americano teve P-valor = 0,427).

12. O mais abrangente deles é o trabalho de Flores, A. W., Bechtel, K. e Lowenkamp, C. T. 2016. “False Positives, False Negatives, and False Analyses: A Rejoinder to Machine Bias: There’s Software Used across the Country to Predict Future Criminals. And It’s Biased against Blacks.” [Falsos positivos, falsos negativos e análises falsas: uma réplica à parcialidade de máquina: Existe um programa de computador no país para prever criminosos futuros. E ele é tendencioso contra negros]. *Fed. Probation* [Diário de condenação federal] 80: 38.

13. Arrow, K. J. 1950. “A difficulty in the concept of social welfare” [Uma dificuldade no conceito de previdência social]. *Journal of political economy* [Revista de economia política] 58, nº 4: 328–46.

14. Young, H. P. 1995. *Equity: in theory and practice* [Equidade: teoria e prática]. Princeton University Press.

15. Dwork, C., Hardt, M., Pitassi, T., Reingold, O. e Zemel, R. 2012. “Fairness through awareness” [Justiça através da consciência]. Nos *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference* [Anais da 3ª Conferência de Inovações em Ciência da Computação Teórica], p. 214–26. ACM.

Capítulo 7: Os alquimistas dos dados

1. Posso sugerir um livro excelente se você quiser aprender mais sobre expectativa de gol e outros algoritmos no futebol. É chamado *Soccermatics* [A matemática do futebol].

2. Há algumas sutilezas aqui. Para fazer essa comparação, vamos supor que colocamos os especialistas de futebol para assistir a um jogo até o momento exato em que a tentativa de gol foi feita, mas não mostramos para eles se foi gol ou não. Perguntamos a eles, então, se a tentativa foi uma grande oportunidade ou não, e consideramos a avaliação deles uma previsão. Tentativas rotuladas de grandes oportunidades são previsões de gols, enquanto aquelas que não são rotuladas de grandes oportunidades são previsões de bolas para fora ou bolas defendidas. Nas narrações de futebol, essas

interpretações de grandes oportunidades correspondem aos pronunciamentos frequentemente usados pelos comentaristas de que um jogador “deveria ter marcado dali”.

3. Julia não teve acesso ao algoritmo Compas. Então, ela usou a engenharia reversa nele. Os níveis de acurácia que encontrou no algoritmo reverso normalmente replicaram os níveis declarados pela Northpointe em outros estudos.

4. A fim de comparar o nível de acurácia de meu modelo (baseado puramente em idade e antecedentes) com o modelo Compas (baseado em um monte de métricas e respostas de um questionário), usei o teste configurado para obter uma curva ROC. A AUC do meu modelo foi de 0,733, que é comparável à capacidade de previsão do modelo Compas. Em outras palavras, um modelo simples representado por idade e antecedentes é (para o banco de dados do Condado de Broward) tão acurado quanto o modelo Compas.

Capítulo 8: Nate Silver *versus* o restante de nós

1. www.yougov.co.uk/news/2017/06/09/the-day-after
2. www.nytimes.com/2016/11/01/us/politics/hillary-clinton-campaign.html?mcubz=3
3. www.nytimes.com/2016/11/10/technology/the-data-said-clinton-would-win-why-you-shouldnt-have-believed-it
4. www.fivethirtyeight.com/features/the-media-has-a-probability-problem
5. www.cafe.com/carl-digglers-super-tuesday-special
6. Tetlock, P. E. e Gardner D. 2016. *Superprevisões: A arte e a ciência de antecipar o futuro*. Objetiva.
7. Esse número foi tirado de um projeto do The Data Face [A cara dos dados] baseado nos dados coletados pelo Projeto do Bom Julgamento: www.thedataface.com/good-judgment-open-election-2016
8. PredictIt foi usado em apenas uma eleição presidencial, mas um de seus precursores que funciona com os mesmos princípios, Intrade, foi usado em ambas as votações de 2008 e 2012. Alex e eu pegamos essas previsões de probabilidade final para cada estado nas três eleições e as comparamos com as previsões do FiveThirtyEight.
9. Rothschild, D. 2009. “Forecasting elections: Comparing prediction markets, polls, and their biases” [Prevedendo eleições: Comparando mercados previsores, pesquisas de opinião e suas parcialidades]. *Public Opinion Quarterly* [Opinião Pública Trimestral] 73, nº 5: 895–916.

10. A pontuação de Brier é um único número que leva em consideração tanto a acurácia quanto a coragem das previsões. Essa pontuação é calculada por meio do quadrado da distância entre a previsão e o resultado verdadeiro. Então, se p é a previsão, dada como uma probabilidade, e σ é o resultado, em que $\sigma=1$, se o evento ocorreu, e $\sigma=0$ se o evento não ocorreu, então $(p - \sigma)^2$ é a pontuação Brier para o evento. Quanto menor a pontuação Brier, melhor. Uma previsão muito corajosa de que um evento ocorrerá com 100% de certeza, que provou ser verdadeira, tem a melhor pontuação Brier possível, 0. Por outro lado, se essa previsão corajosa estiver errada, ele recebe a pior pontuação Brier, 1. A previsão covarde de um evento ter uma chance de 50% de acontecer recebe uma pontuação de Brier de 0,25, independentemente do resultado.

11. Rothschild, D. 2009. “Forecasting elections: Comparing prediction markets, polls, and their biases” [Prevendo eleições: Comparando mercados previsores, pesquisas de opinião e suas parcialidades]. *Public Opinion Quarterly* [Opinião Pública Trimestral] 73, nº 5: 895–916.

12. A maior parte dos estudos empíricos diz respeito à performance da diretoria da empresa, uma tarefa intimamente relacionada com a previsão de negócios futuros. Por exemplo, veja: Ali, M., Lu Ng, Y. e Kulik, C. T. 2014. “Board age and gender diversity: A test of competing linear and curvilinear predictions” [Idade da diretoria e diversidade de gênero: Um teste de previsões competitivas lineares e curvilíneas]. *Journal of Business Ethics* [Revista de Ética em Negócios] 125, nº 3: 497–512. Para uma revisão geral sobre esse assunto, veja Page, S. E. 2010. *Diversity e complexity* [Diversidade e complexidade]. Princeton University Press.

13. Para mais informações sobre as dificuldades em derrotar uma multidão inteligente, veja: Buchdahl, J. 2016. *Squares & Sharps, Suckers & Sharks: The Science, Psychology & Philosophy of Gambling* [Quadrados e afiados, otários e monstros: Ciência, psicologia & filosofia do jogo de apostas]. Vol. 16. Oldcastle Books.

Capítulo 9: Nós “também gostamos” da internet

1. Esse modelo é normalmente chamado “apego preferencial” na literatura matemática, mas possui uma variedade de nomes refletindo a variedade de vezes que foi descoberto. A melhor descrição matemática sobre como ele é usado e funciona pode ser encontrada no artigo de Mark Newman sobre leis de potência: Newman, M. E. J. 2005. “Power laws, Pareto distributions and Zipf’s law” [Leis de potência, distribuição de Pareto e lei de Zipf]. *Contemporary Physics* [Física Contemporânea] 46, nº 5: 323–51.

2. Uma descrição detalhada do modelo “também gostaram” é dada a seguir. Em cada etapa do modelo, um novo cliente chega e analisa o site de seu autor preferido. A probabilidade de que um autor em particular, i , seja o preferido do novo cliente depende das vendas anteriores e é dada por:

$$\frac{n_i + 1}{N + 25}$$

em que n_i é o número de livros vendidos pelo autor i e $N = \sum_{i=1}^{25} n_i$ é o número total de livros vendidos por todos os autores. Uma vez no site de seu autor preferido, o cliente decide comprar um livro novo baseado no número de “também gostaram” compras comuns entre seu autor preferido e outros autores. Então, a probabilidade que ele compre um livro do autor j é

$$\frac{c_{ij} + 1}{n_i + 25}$$

em que c_{ij} é o número de livros vendidos pelo autor i por meio de uma conexão com o autor j . Depois aumentamos c_{ij} , c_{ji} e n_i em uma unidade para consolidar a nova compra, e o próximo cliente chega. Na Figura 9.1 (a) e (b) o raio do círculo próximo ao autor é proporcional a n_i e a espessura do link entre os autores é dada por c_{ij} .

3. Lerman, K. e Hogg, T. 2010. “Using a model of social dynamics to predict popularity of news” [Usando um modelo de dinâmica social para prever popularidade de notícias]. *Proceedings of the 19th international conference on World Wide Web* [Anais da 19ª Conferência Internacional sobre Internet], pp. 621–30. ACM.
4. Burghardt, K., Alsina, E. F., Girvan, M., Rand, W. e Lerman, K. 2017. “The myopia of crowds: Cognitive load and collective evaluation of answers on Stack Exchange” [A miopia das multidões: Carga cognitiva e avaliação coletiva das respostas na Stack Exchange]. *PloS one* 12, nº: e0173610
5. Salganik, M. J., Dodds, P. S. e Watts, D. J. 2006. “Experimental study of inequality and unpredictability in an artificial cultural market” [Estudo experimental de desigualdade e imprevisibilidade num mercado cultural artificial]. *Science* 311, nº 5762: 854-6. Salganik, M. J. e Watts, D. J. 2008. “Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market” [Tirando do rumo: Um estudo experimental de profecias realizadas em um mercado cultural artificial]. *Social psychology quarterly* [Psicologia social trimestral] 71, nº 4: 338–55.
6. Muchnik, L., Aral, S. e Taylor, S. J. 2013. “Social influence bias: A randomized experiment” [Influência social tendenciosa: Um experimento aleatório]. *Science* 341, nº 6146: 647–51.
7. www.popularmechanics.com/science/health/a9335/upvotes-downvotes-and-the-science-of-the-reddit-hivemind-15784871

8. O algoritmo PageRank funciona supondo que navegamos na internet seguindo links aleatoriamente. A página com a classificação mais alta é a mais visitada como resultado dessa navegação aleatória. Para uma descrição matemática do algoritmo do Google, veja Franceschet, Massimo. 2011. “PageRank: Standing on the shoulders of giants” [PageRank: Sobre os ombros do gigante]. *Communications of the ACM* [Comunicações da ACM] 54, nº 6: 92–101.

9. Essa hipótese não foi testada explicitamente, mas estudos mostrando que a nota média no Goodreads decresce quando livros receberam prêmios respeitados corroboram essa hipótese. Veja Kovács, B. e Sharkey, A. J. 2014. “The paradox of publicity: How awards can negatively affect the evaluation of quality” [O paradoxo da publicidade: Como prêmios podem afetar negativamente a avaliação de qualidade]. *Administrative Science Quarterly* [Ciência Administrativa Trimestral] 59, nº 1: 1–33.

Capítulo 10: O concurso de popularidade

1. Giles, J. 2005. “Science in the web age: Start your engines” [Ciência na era da internet: Liguem os motores]. *Nature* 438 (7068), 554–5.

2. www.backchannel.com/the-gentleman-who-made-scholar-d71289d9a82d#.ld8ob7qo9

3. Eom, Y-H. e Fortunato, S. 2011. “Characterizing and modeling citation dynamics” [Caracterizando e modelando citações dinâmicas]. *PloS one* 6, nº 9: e24926.

4. Analisemos isso matematicamente. O gráfico mostra a probabilidade, p , de um artigo ser citado mais de n vezes ou mais. Como os dados são representados numa escala logarítmica e existe uma relação entre eles, temos que

$$\log(p) = \log(k) - a \log(n)$$

em que a é a inclinação da reta e k é uma constante. Daí podemos concluir que

$$p = kn^{-a}$$

que a relação da lei de potência.

5. May, R. M. 1997. “The scientific wealth of nations” [A saúde científica das nações]. *Science*, 275, nº 5301: 793–6.

6. Petersen, A. M., et al. 2014. “Reputation and impact in academic careers” [Reputação e impacto em carreiras científicas]. *Proceedings of the National Academy of Sciences* [Anais da Academia Nacional de Ciências] 111, nº 43: 15316–21.
7. Higginson, A. D. e Munafò, M. R. 2016. “Current incentives for scientists lead to underpowered studies with erroneous conclusions” [Incentivos atuais para cientistas resultam em estudos fracos com conclusões errôneas]. *PLoS biology* . 14: 11, e2000995.
8. Penner, O., Pan, R. K., Petersen, A. M., Kaski, K. e Fortunato, S. 2013. “On the predictability of future impact in science” [Sobre a previsibilidade de impacto futuro na ciência]. *Scientific Reports* [Relatórios científicos] 3.
9. Acuna, D. E., Allesina, S. e Kording, K. P. 2012. “Future impact: Predicting scientific success” [Impacto futuro: Prevendo sucesso científico]. *Nature* 489, nº 7415: 201–2.
10. Sinatra, R., Wang, D., Deville, P., Song, C. e Barabási, A-L. 2016. “Quantifying the evolution of individual scientific impact” [Quantificando a evolução de impacto científico individual]. *Science* 354, nº 6312: aaf5239.
11. Tyson, G., Perta, V. C., Haddadi, H. e Seto, M. C. 2016. “A first look at user activity on Tinder” [Uma impressão inicial da atividade de um usuário do Tinder]. *Advances in Social Networks Analysis and Mining (Asonam)* [Avanços em Análise de Redes Sociais e Pesquisa de Dados], 2016 IEEE/ACM International Conference p. 461-6. IEEE
12. www.collective-behavior.com/the-unstable-dating-game

Capítulo 11: Borbulhando

1. Neste livro e no TED Talk sobre bolhas-filtro, Eli Pariser revelou o quanto nossas atividades on-line são personalizadas. Pariser, Eli. 2011. *The Filter Bubble: How the new personalized web is changing what we read and how we think* [A bolha-filtro: Como a nova internet personalizada está mudando o que lemos e como pensamos]. Penguin. Google, Facebook e outras companhias da internet armazenam dados documentando as escolhas que fazemos quando navegamos on-line e, então, os usam para decidir o que nos mostrar no futuro.
2. <https://newsroom.fb.com/news/2016/04/news-feed-fyi-from-f8-how-news-feed-works>
3. www.techcrunch.com/2016/09/06/ultimate-guide-to-the-news-feed
4. No modelo, a probabilidade de um usuário escolher o *Guardian* no tempo t é igual a

$$\frac{G(t)^2 + K^2}{G(t)^2 + K^2 + T(t)^2 + K^2}$$

em que $G(t)$ é o número de vezes que o usuário já escolheu o *Guardian* e $T(t)$ é o número de vezes que o usuário escolheu o *Telegraph*. Os termos ao quadrado capturam o feedback combinado de (seu interesse no jornal) $\times$ (proximidade ao amigo que compartilhou o artigo). Na simulação mostrada na figura, $K = 5$.

5. Del Vicario, M., et al. 2016. “The spreading of misinformation on-line” [A difusão on-line de informações enganosas]. *Proceedings of the National Academy of Sciences* [Anais da Academia Nacional de Ciências] 113, nº 3: 554–9.

6. www.pewresearch.org/fact-tank/2014/02/03/6-new-facts-about-facebook

7. Os cientistas também realizaram um experimento controlado em que nenhuma mensagem era mostrada. O resultado em termos de número de votos foi muito parecido com a mensagem sem as fotos dos amigos, sugerindo que mensagens não sociais não são particularmente úteis. Todos os detalhes podem ser encontrados em: Bond, R. M., et al. 2012. “A 61-million-person experiment in social influence and political mobilization” [Um experimento com 61 milhões de pessoas sobre influência social e mobilização política]. *Nature* 489, nº 7415: 295–8.

Capítulo 12: O futebol importa

1. Uma série de artigos de Huckfeldt revela uma (para mim) surpreendente diversidade em discussões políticas. Huckfeldt, R., Beck, P. A., Dalton, R. J., & Levine, J. 1995. “Political environments, cohesive social groups, and the communication of public opinion” [Ambiente político, grupos sociais coesivos e a comunicação da opinião pública]. *American Journal of Political Science* [Revista Americana de Ciência Política], 1025–54.

2. Huckfeldt, R. 2017. “The 2016 Ithiel de Sola Pool Lecture: Interdependence, Communication, and Aggregation: Transforming Voters into Electorates” [A Aula Ithiel de Sola Pool de 2016: Interdependência, Comunicação e Agregação: Transformando eleitores em eleitorado]. *PS: Political Science & Politics* [Ciência Política & Política] 50, nº 1: 3–11.

3. DiFranzo, D. e Gloria-Garcia, K. 2017. “Filter bubbles e fake news” [Bolhas-filtro e notícias falsas]. *XRDS: Crossroads, The ACM Magazine for Students* [A Revista ACM para Estudantes] 23, nº 3: 32–5.

4. Jackson, D., Thorsen, E. e Wring, D. 2016. “EU Referendum Analysis 2016: Media, Voters and the Campaign” [Análise de referendun de 2016 da UE: Mídia, eleitores e a campanha].

5. O número seis não deve ser considerado literalmente. As medidas mais recentes do Facebook resultam um grau médio de separação de 4,57. No Twitter, os graus de separação são ainda menores.
6. Johansson, J. 2017. “A Quantitative Study of Social Media Echo Chambers” [Um estudo quantitativo de mídia social em câmaras de eco], tese de mestrado, Universidade de Uppsala.
7. Esses resultados foram tirados de Kulshrestha, J., Eslami, M., Messias, J., Zafar, M. B., Ghosh, S., Gummadi, K. P. e Karahalios, K. 2017. “Quantifying search bias: Investigating sources of bias for political searches in social media” [Quantificando busca tendenciosa: Investigando fontes de parcialidade para buscas políticas na mídia social]. *arXivpreprint arXiv:1704.01347*.

Capítulo 13: Quem lê fake news?

1. www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm_term=.rrw0PaV3wP#.qr7rjqJeAj
2. Allcott, H. e Gentzkow, M. 2017. “Social media and fake news in the 2016 election” [Mídia social e notícias falsas na eleição de 2016]. Nº w23089. National Bureau of Economic Research [Bureau Nacional de Pesquisas Econômicas].
3. Para mais detalhes, veja: www.washingtonpost.com/news/the-fix/wp/2016/11/14/googles-top-news-link-for-final-election-results-goes-to-a-fake-news-site-with-false-numbers
4. De acordo com o SharedCount, em janeiro de 2017, havia 530.858 links na página do Facebook.
5. Franks, N. R., Gomez, N., Goss, S. e Deneubourg, J.-L. 1991. “The blind leading the blind in army ant raid patterns: testing a model of self-organization” (*Hymenoptera: Formicidae*) [O cego liderando o cego em padrões de ataque de formigas guerreiras: testando um modelo de auto-organização]. *Journal of Insect Behavior* [Revista do Comportamento de Insetos] 4, nº 5: 583–607.
6. Ward, A. J. W., Herbert-Read, J. E., Sumpter, D. J.T. e Krause, J. 2011. “Fast and accurate decisions through collective vigilance in fish shoals” [Decisões rápidas e acuradas através da vigilância coletiva em cardumes de peixes]. *Proceedings of the National Academy of Sciences* [Anais da Academia Nacional de Ciências] 108, nº 6: 2312–5.
7. Biro, D., Sumpter, D. J. T., Meade, J. e Guilford, T. 2006. “From compromise to leadership in pigeon homing” [De concessão à liderança no retorno de pombos para casa]. *Current Biology* [Biologia Atual] 16, nº 21: 2123–8.

8. Cassino, D. e Jenkins, K. 2013. “Conspiracy Theories Prosper: 25 per cent of Americans Are ‘Truthers’” [Teorias da Conspiração prosperam: 25% dos americanos “procuram a verdade sobre os acontecimentos de 11/9”]. Boletim de Imprensa — <http://publicmind.fdu.edu/2013/outthere>.
9. Veja, por exemplo, essa resposta do Facebook. <http://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news>
10. Loader, B. D., Vromen, A. e Xenos, M. A. 2014. “The networked young citizen: social media, political participation and civic engagement” [Os cidadãos jovens conectados: mídia social, participação política e engajamento civil]: 143–50.

Capítulo 14: Aprendendo a ser sexista

1. Existe uma vasta gama de observações e experimentos confirmando essa parcialidade. Por exemplo, em seu livro de 2003 *The American Non-Dilemma: Racial inequality without racism* [O não dilema americano: Desigualdade racial sem racismo], Nancy DiTomaso documenta um favoritismo intragrupo extensivo entre americanos brancos.
2. Lavergne, M. e Mullainathan, S. 2004. “Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination” [São Emily e Greg mais empregáveis do que Lakisha e Jamal? Um experimento de campo sobre discriminação no trabalho]. *The American Economic Review* [Revista Americana de Economia] 94, nº 4: 991–1013.
3. Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J. e Handelsman, J. 2012. “Science faculty’s subtle gender biases favor male students” [A leve parcialidade de gênero dos docentes em ciências favorece estudantes do sexo masculino]. *Proceedings of the National Academy of Sciences* [Anais da Academia Nacional de Ciências] 109, nº 41: 16474–9.
4. Greenwald, A. G., McGhee, D. E. e Schwartz, J. L. K. 1998. “Measuring individual differences in implicit cognition: the implicit association test” [Medindo diferenças individuais na cognição implícita: o teste de associação implícita]. *Journal of Personality and Social Psychology* [Revista da Personalidade e Psicologia Social] 74, nº 6: 1464.
5. Faça o teste você mesmo em www.implicit.harvard.edu/implicit/takeatest
6. Resposta: um boi.
7. www.theguardian.com/technology/2016/dec/04/googledemocracy-truth-internet-search-facebook
8. www.theguardian.com/technology/2016/dec/05/google-alternates-search-autocomplete-remove-are-jews-evil-suggestion

9. Tosik, M., Hansen, C. L., Goossen, G. e Rotaru, M. 2015. “Word embeddings vs word types for sequence labeling: the curious case of CV parsing” [Semântica distribucional *versus* tipos de palavras para rotular sequências: o caso curioso de análise de CV]. In *VS@ HLT-NAACL*, p. 123–8.
10. Bolukbasi, T., Chang, K-W, Zou, J. Y., Saligrama, V. e Kalai, A. T. 2016. “Man is to computer programmer as woman is to homemaker? Debiasing word embeddings” [Homem está para programação de computadores assim como mulher está para dona de casa? Tirando a parcialidade da semântica distribucional]. Em *Advances in Neural Information Processing Systems* [Avanços nos Sistemas de Processamento de Informação Neural], p. 4349–57.
11. Para uma revisão, veja Hofmann, W., Gawronski, B., Gschwendner, T., Le, H. e Schmitt, M. 2005. “A meta-analysis on the correlation between the implicit association test and explicit self-report measures” [Uma meta-análise na correlação entre o teste de associação implícita e medidas de autoavaliação explícitas]. *Personality and Social Psychology Bulletin* [Boletim de Personalidade e Psicologia Social] 31, nº 10: 1369–85.

Capítulo 15: Apenas pensamento entre decimal

1. Uma porta parecida pode ser encontrada em computação quântica e é conhecida como porta de Hadamard.
2. Para implementar esse modelo, usei o programa demo de redes neurais em <https://lecture-demo.ira.uka.de>. Com ele você pode brincar com uma variedade de ferramentas para redes neurais.
3. www.tensorflow.org
4. Duas descrições boas de redes neurais podem ser encontradas em <http://colah.github.io/posts/2015-08-Understanding-LSTMs> e <http://karpathy.github.io/2015/05/21/rnneffectiveness>
5. Vinyals, O. e Le, Q. 2015. “A neural conversational model” [Um modelo neural conversacional]. *arXiv preprint arXiv:1506.05869*.
6. Sutskever, I., Vinyals, O. e Le, Q. 2014. “Sequence to sequence learning with neural networks” [Aprendizado seq2seq com redes neurais]. *Advances in Neural Information Processing Systems* [Avanços em Sistema de Processamento de Informação Neural], p. 3104–12.
7. Mikolov, T., Joulin, A. e Baroni, M. 2015. “A roadmap towards machine intelligence” [Um mapa para a inteligência de máquina]. *arXiv preprint arXiv:1511.08130*.
8. Também tomei a liberdade de corrigir erros de pontuação.

Capítulo 16: Arrase no *Space Invaders*

1. <http://static.ijcai.org/proceedings-2017/0772.pdf>
2. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., *et al.* . 2015. “Human-level control through deep reinforcement learning” [Controle em nível humano através de aprendizagem reforçada]. *Nature* 518, nº 7540: 529–33.
3. O conjunto de dados usado está descrito aqui: www.image-net.org/about-overview
4. www.qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world
5. Detalhes sobre as redes usadas e os ganhadores podem ser encontrados aqui: <http://cs231n.github.io/convolutional-networks/#case>
6. <http://selfdrivingcars.mit.edu>
7. <https://arxiv.org/pdf/1609.08144.pdf>
8. <https://arxiv.org/pdf/1708.04782.pdf>

Capítulo 17: O cérebro bacteriano

1. www.youtube.com/watch?v=OFBwz4R6Fi0&feature=youtu.be
2. www.haggstrom.blogspot.se/2013/10/guest-post-by-david-sumpter-why
3. Häggström, O. 2016. *Here Be Dragons: Science, Technology and the Future of Humanity* [Aqui há dragões: ciência, tecnologia e o futuro da humanidade]. Oxford University Press.
4. www.youtube.com/watch?v=V0aXMTpZTfc
5. Hassabis, D., Kumaran, D., Summerfield, C. e Botvinick, M. 2017. “Neuroscience-inspired artificial intelligence” [Neurociência inspirada por inteligência artificial]. *Neuron* 95, nº 2: 245–58.
6. Kaminski, J. e Nitzschner, M. 2013. “Do dogs get the point? A review of dog–human communication ability” [Os cachorros entendem? Uma análise da habilidade de comunicação entre cães e humanos]. *Learning and Motivation* [Aprendizado e Motivação] 44, nº 4: 294–302.
7. O texto que segue é baseado no resumo Chittka, Lars. 2017. “Bee cognition” [Cognição da abelha]. *Current Biology* [Biologia Atual] 27, nº 19: R1049–53.

8. Loukola, O. J., Clint P. J., Coscos, L., e Chittka, Lars. “Bumblebees show cognitive flexibility by improving on an observed complex behavior” [Abelhas grandes mostram flexibilidade cognitiva ao se aperfeiçoarem ao observar comportamento complexo]. *Science* 355, nº 6327 (2017): 833–6.
9. Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., Vitale, L., et al. 2013. “An estimation of the number of cells in the human body” [Uma estimativa do número de células do corpo humano]. *Annals of Human Biology* [Anais da Biologia Humana] 40, nº 6: 463–71.
10. Herculano-Houzel, S. 2009. “The human brain in numbers: a linearly scaled-up primate brain” [O cérebro humano em números: um cérebro primata ampliado linearmente]. *Frontiers in Human Neuroscience* [Fronteiras em Neurociência Humana], vol. 3.
11. Scholz, M., Dinner, A. R., Levine, E. and Biron, D. 2017. “Stochastic feeding dynamics arise from the need for information and energy” [Dinâmicas de alimentação estocástica surgem da necessidade por informação e energia]. *Proceedings of the National Academy of Sciences* [Anais da Academia Nacional de Ciências] 114, nº 35: 9261–6.
12. Boisseau, R. P., Vogel, D. e Dussutour, A. 2016. “Habituation in non-neural organisms: evidence from slime moulds” [Habituação em organismos não neurais: evidência a partir de bolor limoso]. In *Proc. R. Soc. B*, vol. 283, nº 1829, p. 20160446. The Royal Society [Sociedade Real].
13. Analiso isso com mais detalhes no artigo: Ma, Q., Johansson, A., Tero, A., Nakagaki, T. and Sumpter, D. J. T. 2013. “Current-reinforced random walks for constructing transport networks” [Passeios aleatórios de corrente reforçada na construção de redes de transporte]. *Journal of the Royal Society Interface* [Revista da Interface da Sociedade Real] 10, nº 80: 20120864.
14. Baker, M. D. e Stock, J. B. 2007. “Signal transduction: networks and integrated circuits in bacterial cognition” [Transdução de sinal: redes e circuitos integrados na cognição de abelhas]. *Current Biology* [Biologia Atual] 17, nº 23: R1021–4.
15. Turing, A. M. 1950. “Computing machinery and intelligence” [Maquinaria de computação e inteligência]. *Mind* 59, nº 236: 433–60.
16. Pesquisei um exemplo desse tipo no seguinte artigo: Herbert-Read, J. E., Romenskyy, M. e Sumpter, D. J. T. 2015. “A Turing test for collective motion” [Um teste de Turing para movimentação coletiva]. *Biology letters* 11, nº 12: 20150674.
17. www.facebook.com/zuck/posts/10154361492931634

Capítulo 18: De volta à realidade

1. Apesar de você conseguir achar isso no Reddit, é claro:
[www.reddit.com/r/TheSilphRoad/comments/6ryd6e/cumulative_probability_legendary_raid_boss_ca](http://www.reddit.com/r/TheSilphRoad/comments/6ryd6e/cumulative_probability_legendary_raid_boss_catch)
tch

Agradecimentos

Obrigado a todas as pessoas que entrevistei ou que responderam às minhas perguntas por e-mail para este livro. Elas incluem, mas não estão limitadas a, Adam Calhoun, Alex Kogan, Amit Datta, Angela Grammatas, Anupam Datta, Bill Dieterich, Bob Huckfeldt, Brian Connelly, “CCTV Simon”, David Silver, Emilio Ferrara, Garry Gellade, Glenn McDonald, Harm van Seijen, Hunt Allcott, Joanna Bryson, Johan Ydring, Julia Dressel, Katie Genter, Kathleen Richardson, Kristina Lerman, Marc Keuschnigg, Matthew Gentzkow, Michael Wood, Michela Del Vicario, Mona Chalabi, Olle Häggström, Oriol Vinyals, Santo Fortunato, Tassos Noulas, Tim Brennan, Tim Laue e Tomas Mikolov. Aprendi muito com todos vocês. Muito obrigado.

Um obrigado especial a Renaud Lambiotte e Michal Kosinski pelas longas conversas que tive com ambos assim que comecei este livro. Vocês me forneceram a estrutura racional adequada para que eu pudesse ponderar sobre os perigos, reais e imaginários, dos algoritmos. Este livro não seria nem de longe tão bom sem tal contribuição.

Agradeço à mamãe e ao papai por lerem este livro tão meticulosamente e por todos os comentários. Mesmo que agora eu já seja um adulto, ainda sinto que vocês me apoiam bastante.

Na Bloomsbury, gostaria de agradecer a Rebecca Thorne a viagem ao Google que iniciou este livro: a Anna MacDiarmid, por toda sua contribuição e encorajamento, e a Jim Martin, por sua fé em mim. Ainda

não sei como transformar este livro em uma desculpa para ver os Pars juntos, mas pensarei em algo.

Obrigado a Emily Kearns pela revisão cuidadosa e inúmeras melhorias feitas no texto.

Um agradecimento especial a Alex Szorkovszky pelo seu trabalho com o gerador Tolstói, medindo a acurácia das eleições presidenciais, e na dissecação do Tinder. Obrigado também por aguentar minhas oscilações longas e incontrolláveis sobre os prós e contras de algoritmos. Esta última parte de agradecimento precisa ser também estendida ao meu grupo de pesquisa em Uppsala. Obrigado, em particular, aos meus estudantes de doutorado (Ernest, Björn e Linnea), que foram tão pacientes comigo durante a escrita deste livro.

Obrigado a Anna Baddeley por se encarregar dos artigos sobre *Dominados pelos números* para a *Economist* 1843, e dar o retorno sobre eles. Aprendi bastante sobre o nível e o equilíbrio necessários ao se escrever sobre matemática.

Agradeço a Joakim Johansson por fazer seu projeto de mestrado sobre o Twitter e me ajudar a compreender a câmara de eco nessa plataforma.

A ideia de *Dominados pelos números* evoluiu a partir de conversas (no dia seguinte a minha visita ao Google) com meu agente Chris Wellbelove sobre escrever um livro chamado *Maths is Evil* [A matemática é perversa]. A matemática não se mostrou ser assim tão perversa quanto talvez a tenhamos visto naquele dia, mas, como sempre, sou grato pela sua maneira clara de raciocínio. Aprendi muito com você.

O título propriamente dito é devido a minha esposa Lovisa, que o falou para mim, reflexivamente, quando lhe disse sobre o que gostaria de escrever. Obrigado por isso e muito mais. Desfruto todos os dias da minha vida com você.

Este livro não teria sido possível sem a ajuda dos meus filhos, Elise e Henry. Claramente, vocês dois compreendem tudo relacionado a nossa vida on-line muito melhor do que eu, então, obrigado por me explicá-la tão pacientemente. E por serem as melhores crianças do mundo.

Índice

70 News
abelhas
Acharya, Anurag
Adamic, Lada
Agência de Projetos de Pesquisa Avançada de Defesa (Darpa) EUA
agregadores de notícias
Albright, Jonathan
aleatoriedade
algoritmos
 algoritmo Compas
 algoritmo PCRA
 algoritmos da caixa preta
 AlphaGo Zero
 Amazon
 análise de personalidade
 calibração
 eliminando parcialidade
 filtro de algoritmos
 Google
 Libratus
 linguagem
 pesquisas preditivas
 prevendo resultados de futebol
 redes neurais
 “também gostaram”
 Word2vec
Allcott, Hunt
Amazon
análise de componente principal (ACP)
 algoritmo Compas
 classificando personalidades
análise de eleitores

análise de personalidade
Big Five
analogia de palavras
Angwin, Julia
Apple
Apple Music
Aral, Sinan
Arrow, Kenneth
ASI Data Science
Atari

bactéria (*E. coli*)
bancos de dados
 projeto myPersonality
Banksy
 “Banksy islâmico”
Bannon, Steve
Barabási, Albert-László
BBC
 BBC Bitesize
Bezos, Jeff
Biederman, Felix
Biro, Dora
BlackHatWorld
Blizzard
blogs políticos
bolhas-filtro
bolors limosos (*Physarum polycephalum*)
Bolukbasi, Tolga
Bostrom, Nick
Bots
Boxing
Breakout
Breitbart
Brennan, Tim
Brexit
 análise de eleitores
Broome, Fiona
Bryson, Joanna
Buolamwini, Joy
Burrell, Jenna
Bush, George W.
Business Insider
BuzzFeed

cachorros
Cadwalladr, Carole
CAFE
Cambridge Analytica (CA)
 modelos de regressão
câmaras de eco
Cameron, David
Capitão Pugwash
careerchange.com
Chalabi, Mona
chatbots
Chittka, Lars
citações
Clinton, Hillary
CNN
Connelly, Brian
Corbett-Davies, Sam
Cruz, Ted
curvas em forma de sino

dados, veja levantamento de dados on-line
Daily Mail
Daily Star
Datta, Amit
Davis, Steve J.
Deep Blue
Del Vicario, Michela
discussões políticas
distribuição de probabilidade
Dragan, Anca
Dressel, Julia
Drudge Report
DudePerfect
Dugan, Regina
Dussutour, Audrey
Dwork, Cynthia

Economist
Economist
efeito Mandela
Ela (filme)
Eom, Young-Ho

estatística
modelos de regressão
Etzioni, Oren

Facebook
algoritmo de feed de notícias
amigos do Facebook
inteligência artificial (IA)
Messenger
patentes
perfis do Facebook
projeto myPersonality
Will B. Candid

fake news
circuito de feedback
MacronLeaks
mundo pós-verdade

falsos negativos
falsos positivos
Fark
feedback social
Feedly
Feller, Avi
Fergus, Rob
Ferrara, Emilio
FiveThirtyEight
Flipboard
Flynn, Michael
formação de pares românticos
formigas
Fortunato, Santo
Fowler, James
Franks, Nigel
Frostbite
futebol
jogadores robôs

Gates, Bill
Gelade, Garry
Genter, Katie
Gentzkow, Matthew
Geoengineering Watch
Glance, Natalie

GloVe (vetores globais para representação de palavras)

Go

Goel, Sharad

Google

“Não seja perverso”

anúncios personalizados

autopreenchimento do Google

chapéus pretos

DeepMind

Google Acadêmico

Google Notícias

Google+

inteligência artificial (IA)

Pesquisa Google

SharedCount

Gore, Al

gráficos log-log

Grammatas, Angela

Guardian

Guardian US

Häggström, Olle

Here Be Dragons

Hassabis, Demis

Hawking, Stephen

He, Kaiming

Higginson, Andrew

Hinton, Geoffrey

hipótese de singularidade

históricos de busca

históricos de navegação

HotUKDeal

Huckfeldt, Bob

Huffington Post

Independent

índice h

Instagram

inteligência artificial (IA)

limitações

redes neurais

superinteligência

teste de Turing
interferência russa
internet
provedores de serviço de internet (PSIs)
Instituto Allen de Inteligência Artificial
Instituto do Futuro da Humanidade
Intrade
iTunes

Jie, Ke
Johansson, Joakim
Journal of Spatial Science
justiça

Kaminski, Juliane
Kasparov, Garry
Keith, David
Kerry, John
Keuschnigg, Marc
Kleinberg, Jon
Kluemper, Donald
Kogan, Alex
Kosinski, Michal
Kramer, Adam
Krizhevsky, Alex
Kulshrestha, Juhi
Kurzweil, Ray

Lake, Brenden
Laue, Tim
Le Comber, Steve
Le Cun, Yan
Le Pen, Marine
Le, Quoc
Lei do Direito à Moradia dos EUA
leis de potência
Lerman, Kristina
levantamento de dados on-line
análise de componente principal (ACP)
parcialidade de gênero
prevenção
Levin, Simon

Libratus
Liga Nacional de Futebol Feminino (NWSL)
linguagem
LinkedIn
literatura
Luntz, Frank

Macron, Emmanuel
Mandela, Nelson
Martin, Erik
matemática
 avaliação tendenciosa
 leis de potência
 modelos matemáticos
Matrix, The (teoria)
May, Lord Robert
McDonald, Glenn
Medium
Mercer, Robert
Microsoft
Mikolov, Tomas
Minecraft
minhocas (*C. elegans*)
modelos de regressão
Mosseri, Adam
Mrsic-Flogel, Thomas
Ms Pac-Man
Munafò, Marcus
mundo pós-verdade
Musk, Elon

Nature
Natusch, Waffles Pi
Netflix
New York Times
 The Upshot
Nix, Alexander
Northpointe
O'Neil, Cathy
 Weapons of Math Destruction
Noiszy

Obama, Barack

Observer

oferta e procura de emprego

OpenWorm

Overwatch

parcialidade

GloVe (vetores globais para representação de palavras)

justiça e injustiça

parcialidade de gênero

parcialidade de máquina

parcialidade na calibração

parcialidade racial

Partido Conservador

Partido Democrata

Partido Republicano

Partido Trabalhista

Momentum

Pasquale, Frank

The Black Box Society

Paul, Jake

Pennington, Jeffrey

pesquisas de opinião

PewDiePie

Pierson, Emma

Pittsburgh Post-Gazette

PolitiFact

pontuação de Brier

Popular Mechanics

portas lógicas

positivos verdadeiros

Pratt, Stephen

PredictIt

Prince

projeto myPersonality

propaganda

propaganda redirecionada

ProPublica

Pundit

*Q*bert*

Qualtrics

rastros químicos

Reddit

redes neurais

algoritmo Compas

limitações

redes neurais convolucionais

redes neurais recorrentes

reincidência criminal

RiceGum

Richardson, Kathleen

Road Runner

robôs

Robotank

Salganik, Matthew

Sanders, Bernie

Scholz, Monika

Science

SCL

Serviço Nacional de Saúde do Reino Unido (NHS)

serviços de ajuda on-line

Silver, David

Silver, Nate

O sinal e o ruído

Silverman, Craig

Simonyan, Karen

Skeem, Jen

Sky Sports

Snapchat

Snopes

Sopa Campbell

Space Invaders

Spotify

Stack Exchange

StarCraft

Stillwell, David

Sullivan, Andrew

Sumpter, David

Soccermatics

Sun, The

superinteligente

superprevisores

Szorkovszky, Alex

Taleb, Nassim
Tegmark, Max
Telegraph
Telescópio Espacial James Webb
teorias da conspiração
Tesla
Tetlock, Philip
Texas, Virgil
The Gateway
Tidal
Times, The
Tinder
Tolstói, Leo
 Anna Karenina
 Guerra e Paz
trolls
Trump, Donald
 campanha eleitoral
 resultado da eleição
 Twitter
TUI
Turco Mecânico
Turing, Alan
Twitter
 MacronLeaks
Tyson, Gareth

União Europeia (UE)

van Seijen, Harm
Vinyals, Oriol
vloggers

Wall Street Journal
Ward, Ashley
Washington Post
Watts, Duncan
Which?
Wikipédia
Wired
Wood, Mike

Word2vec
Worswick, Steve

Yahoo
Ydring, Johan
YouGov
Young, Peyton
Equity
YouTube
Youyou, Wu

Zuckerberg, Mark

Este e-book foi desenvolvido em formato ePub pela Distribuidora Record de Serviços de Imprensa S.A.

Dominados pelos números

Site do autor:

<http://www.david-sumpter.com/>

Twitter do autor:

<https://twitter.com/soccermatics>

Medium do autor:

<https://medium.com/@Soccermatics>